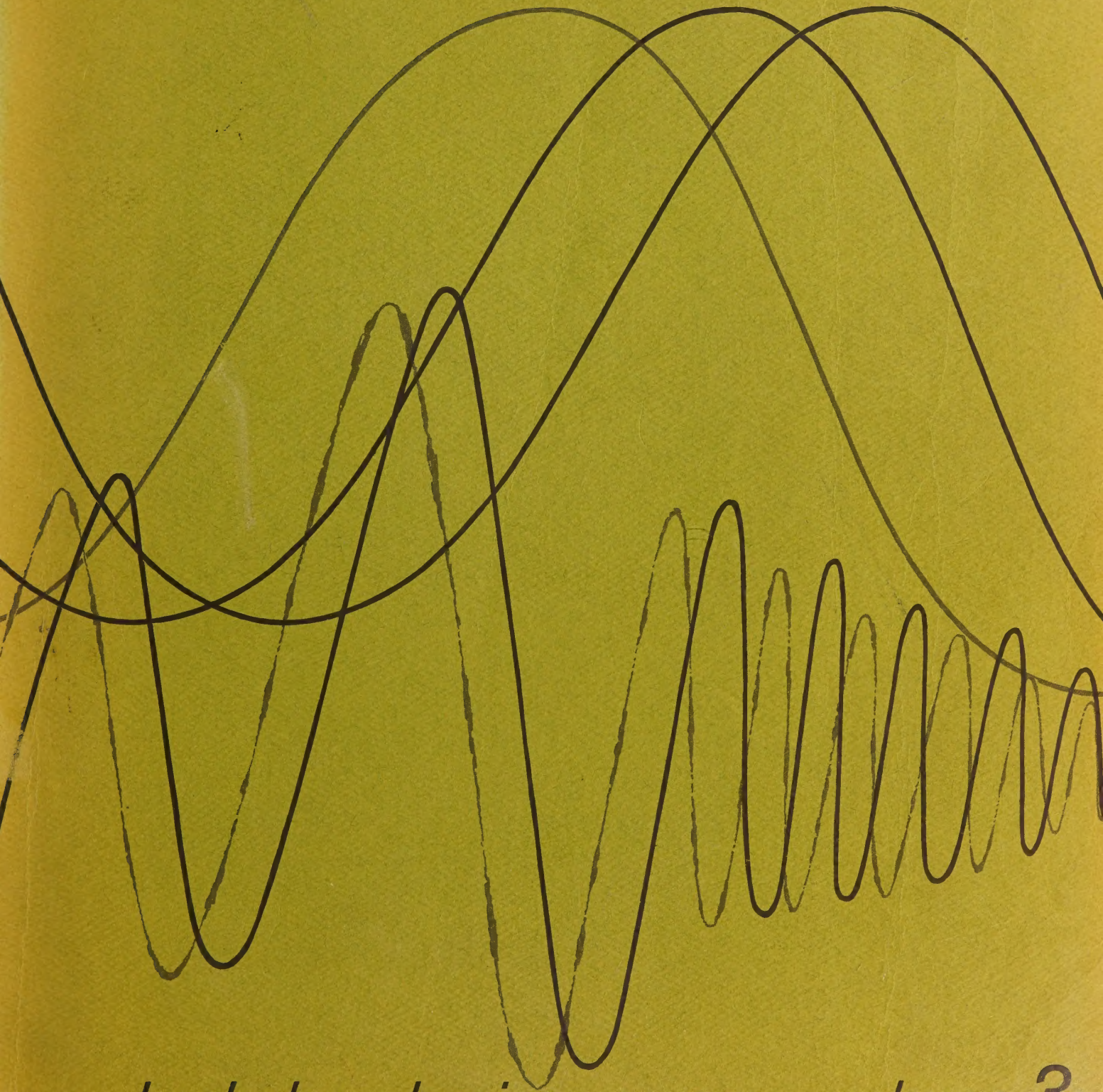
waves and oscillations

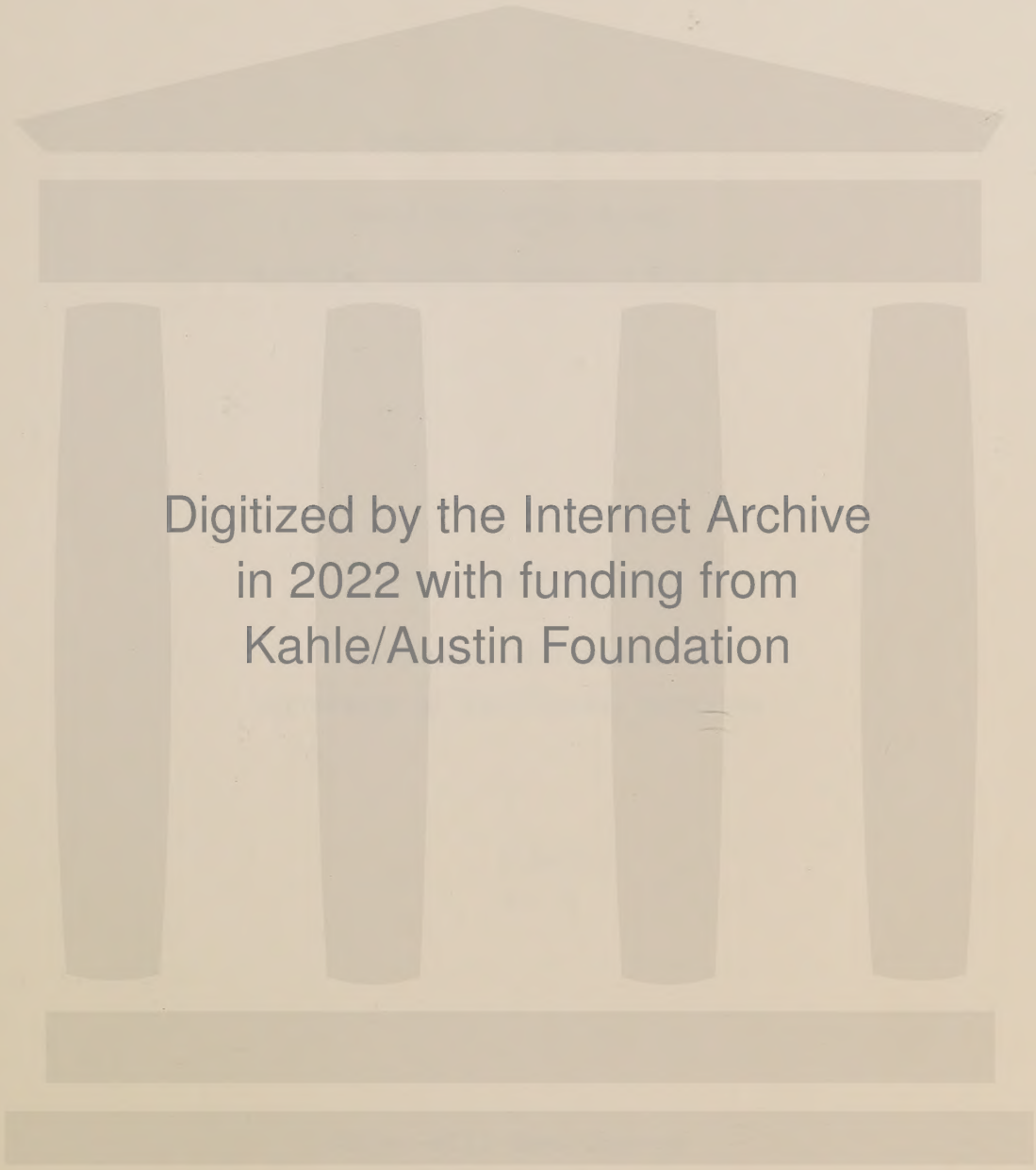
preliminary edition

frank s. crawford, jr.



berkeley physics course—volume 3

ST. JOHN
FISHER
COLLEGE
LIBRARY


Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/wavesoscillation0000unse

Preliminary Edition

WAVES AND OSCILLATIONS

Berkeley Physics Course - Volume 3

FRANK S. CRAWFORD, Jr.

Professor of Physics
University of California, Berkeley

QC
21
.B4445
Vol.3

McGraw-Hill Book Company

New York San Francisco St. Louis Toronto London Sydney

The preparation of this course was supported by a grant from the National Science Foundation to Educational Services Incorporated.

WAVES AND OSCILLATIONS

Copyright © 1964, 1965, 1966 by Educational Services Incorporated. All rights reserved. No part of the material covered by this copyright may be reproduced in any form without the written permission of Educational Services Incorporated.

Preface to Volume III
Preliminary Edition, Fall of 1966

The subject of this volume of the Berkeley Physics Course is waves and oscillations. The material consists of five chapters: (1) Modes; (2) Waves; (3) Emission and absorption of waves: Impedance; (4) Polarization; (5) Interference and diffraction. In addition there is a mathematical appendix and a section of Advanced Topics.

Each of the five chapters consists of about 60 pages of basic text (for a total of about 285 pages), exclusive of certain "starred" paragraphs and sections (totaling 50 pages) and of Problems and Home Experiments (totaling 90 pages). The starred material is not especially difficult, but neither is it essential to the development of the basic material. It can be omitted in a minimal course.

The Problems and Home Experiments should be read, whether or not they are worked or performed. (The answers are given for most of the problems and home experiments.) An Optics Kit is provided for the Home Experiments of Chapters 4 and 5 (on polarization, interference, and diffraction). The kit consists of a diffraction grating, two pieces of polaroid linear polarizer, a quarter-wave plate, a half-wave plate, a circular polarizer, a green filter, a red filter, and two microscope slides. Other components easily obtainable in grocery and hardware stores are not included. The Home Experiments of Chapter 3 include several experiments that require two tuning forks of the same nominal frequency, preferably C523.3 cps, although A440 is all-right. (The reason for preferring C523.3 is that the required cardboard tubes are then only about 12 inches long and can be obtained from many convenient sources -- rolls of paper towels, waxed paper, etc.) The tuning forks are readily available at any music store, and are therefore not included in the kit. I suggest that the teacher read Home Experiments 25 through 33 (Chapter 3) and, if he agrees with me that they are entertaining and instructive, verify with his local music stores that a sufficient number of forks is available.

The introductory section in each chapter contains a brief "preview" of the material to be covered in that chapter. I suggest that the teacher (and student) first survey the entire book by reading these sections (Secs. 1.1, 2.1, 3.1, 4.1, and 5.1). Here is a briefer outline of the book:

Chapter 1 - Modes. We begin by studying several familiar examples of oscillating systems having one degree of freedom, such as a pendulum and an LC circuit. We emphasize that the oscillations are always the result of two opposing tendencies, which may be called "return-force-like" and "inertia-like." (As an interesting unfamiliar example, we consider plasma oscillations in the earth's ionosphere.) In Sec. 1.3 we examine the consequences of our demand that the oscillation amplitudes be sufficiently small so that the "return force" is linearly proportional to the "displacement" (we use quotation marks because we also include non-mechanical examples). One consequence of this linearity is that the "motion" is of the form $A \cos \omega t$. Another consequence is that two or more different motions may be "superposed." In Sec. 1.4 we examine several systems having two degrees of freedom and discover the fact that there are always two independent modes of motion (for two degrees of freedom), each of which is a simple harmonic oscillation of the entire system with a definite configuration or "shape" of the entire system for each mode and a definite frequency (one for each mode). In Sec. 1.5 we detour to study beats between the two modes of two weakly coupled oscillators and between two tuning forks. In Sec. 1.6 we study the modes of systems with more than two degrees of freedom. We learn that there are always exactly as many modes as degrees of freedom; that each mode corresponds to a harmonic vibration of the entire system with a definite frequency ω and a definite spatial configuration; that each moving part i oscillates with a motion $A_i \cos \omega t$, i.e., with its own amplitude A_i, but with a common phase constant for all moving parts; and that the frequency ω and the configuration of amplitudes A_i constitute the complete characteristics of a particular mode. We learn that the most general motion of the system can be regarded as a superposition of the normal modes of the system with a suitable amount and phase constant for each mode. We then go to the limit of a huge number of moving parts in a finite region of space, so that we can treat the system as approximately

"continuous." The modes are then called "standing waves." We find that a single mode or standing wave always has (in common with the motion of a system with one degree of freedom) a single frequency, a physical quantity corresponding to "inertia," and a physical quantity corresponding to "return force." (For a transversely vibrating string, the inertial property is mass density and the "return force" is the transverse component of the string tension.) We also find that if the continuous medium is homogeneous (i.e., if the string, for example, has uniform mass density and tension), then the spatial configuration of the moving parts is always sinusoidal for a given mode. That is, the displacement of the moving parts from equilibrium at a given instant has the form $A \sin (2\pi z/\lambda)$. Thus we learn that in a homogeneous medium a standing wave has the form $A \cos \omega t \sin (2\pi/\lambda)$ for the displacement of a moving part located at equilibrium position z from its equilibrium position. We also learn that the wavelength λ and frequency $\nu = \omega/2\pi$ need not have the simple relation $\lambda\nu = v$, where v is a constant, independent of λ or ν, but that, more generally, v depends on λ (or what is the same thing, on ω). Thus we introduce the idea of a dispersion relationship, which is just the relationship between ν and λ. We find the dispersion relationship for a beaded string and for electromagnetic waves in the earth's atmosphere.

Chapter 2 - Waves. This chapter consists mostly of the study of traveling waves in homogeneous media. We learn that in an "open" system, where the waves emitted by a source are never reflected back so as to impinge again on the source, a steady state of oscillation in which all "moving parts" have motion proportional to $\cos (\omega t + \phi)$ is possible for any frequency ω. (The phase constant ϕ and the amplitude depend on the location of the moving part.) This is in contrast to a "closed" system, where a steady state with motion proportional to $\cos \omega t$ occurs only for the discrete frequencies ω corresponding to the normal modes (the standing waves). We learn that for a given medium (for example, for air of a given mass density and pressure carrying sound waves) the dispersion relation is the same for traveling waves as for standing waves. In Sec. 2.2 we study harmonic traveling waves and learn the famous formula for the phase velocity, $v = \lambda\nu = \omega/k$, where k is the wave number, $2\pi/\lambda$. In Sec. 2.3 we study a superposition of two traveling

waves having almost but not quite the same frequency and discover that the "beats" -- regions where the two waves superpose so as to give a constructive-interference maximum -- do not travel with the phase velocity v, but with a velocity that is called the "group velocity" v_g, with $v_g = d\omega/dk$, where $d\omega/dk$ is obtained from the dispersion relation $\omega = \omega(k)$ and is evaluated at the average wavelength of the two waves. In Sec. 2.4 we study a superposition of many harmonic oscillations of different frequencies each of the form $A \cos \omega t$, with the various ω occupying a band of frequencies of width $\Delta\omega$. We find that the resulting superposition is a function of time $f(t)$ that has the form of a "pulse," i.e., that is substantially different from zero only during a time interval of duration Δt given by $\Delta\nu\, \Delta t \approx 1$. We also study the behavior of a "wave packet" emitted by a transmitter whose oscillations have the form of a pulse. We find that the wave packet or "group" propagates with the group velocity v_g, that the packet has a length Δz related to its band of wave numbers Δk by $\Delta(k/2\pi)\Delta z \approx \Delta\nu\, \Delta t \approx 1$ just after it is emitted. In Sec. 2.5 we study several examples of wave equations, with emphasis on the so-called classical wave equation. We study briefly the three-dimensional wave equation and three-dimensional plane waves having a propagation vector $\underset{\sim}{k} = \hat{x}\, k_x + \hat{y}\, k_y + \hat{z}\, k_z$ such that the components k_x, k_y, and k_z give (at any instant) the number of radians of change of phase per unit distance along x, y, and z. We show that Maxwell's equations for vacuum lead to the three-dimensional classical wave equation for the three components of $\underset{\sim}{E}$ and for the three components of $\underset{\sim}{B}$, with a wave propagation velocity c. We also consider a linear, homogeneous, isotropic medium and show that Maxwell's equations again give the classical wave equation for $\underset{\sim}{E}$ and $\underset{\sim}{B}$, with a wave propagation velocity $v = c/n(\omega)$, where $n(\omega)$ is the index of refraction given by $n(\omega) = [\mu(\omega)\epsilon(\omega)]^{\frac{1}{2}}$, where $\mu(\omega)$ is the magnetic permeability and $\epsilon(\omega)$ is the dielectric constant.

Chapter 3 - Emission and absorption of waves: Impedance. In Chapter 1, the inertia and return force are the two interesting quantities that we use to characterize a medium carrying standing waves. That is because a standing wave is something like a simple harmonic oscillator. In Chapter 3 we find that for traveling waves a more natural (but equivalent) way to characterize

a medium is in terms of two other quantities (closely related to the inertia and return force); these are the propagation velocity v and the "characteristic impedance" Z. Traveling waves carry energy from one place to another, for example from a transmitter to a receiver. The phase velocity v gives the velocity at which the wave energy travels for a single frequency ω; for a superposition of different frequencies it is instead the group velocity v_g that gives the propagation velocity of the energy. The characteristic impedance Z gives the rate at which the transmitter emits energy for a given motion of the transmitter output. We find that the transmitter experiences a "resistive drag" force that opposes its motion with a force proportional to the velocity of the transmitter's output "moving part." The proportionality constant is Z. We find also that we can"terminate" a traveling wave without reflection if we let it impinge on a "lumped" impedance whose value equals the characteristic impedance Z. We give closely parallel discussions of the emission of waves on vibrating string shaken at one end (Sec. 3.2), of sound waves in a tube driven at one end by a piston (Sec. 3.3), of current and voltage waves on a transmission line (Sec. 3.4), and of radiation of electromagnetic plane waves by a huge oscillating sheet of charge (Sec. 3.5). In Sec. 3.6 we use Maxwell's equations to give a general discussion of electromagnetic plane waves in vacuum, without considering specific sources of the waves. The transmission line (Sec. 3.4) and oscillating sheet of charge (Sec. 3.5) are concrete examples illustrating the general results of Sec. 3.6. They may be omitted if necessary. In Sec. 3.7 we consider the total reflection that results when a traveling wave impinges on a reactive medium. Examples are given of reflection of radio waves by the earth's ionosphere, reflection of visible light by a silvered surface, reflection of microwaves in a wave guide beyond cutoff, and total internal reflection of light incident from glass to air at an angle greater than the critical angle. This section can be omitted if necessary.

Chapter 4 - Polarization. The illustrative examples make use of the "slinky" toy and electromagnetic radiation. (There are a large number of Home Experiments using the optics kit. Most of them are quite easy and instructive.) In Sec. 4.2 we discuss the description of transverse waves on a slinky and transverse light waves in terms of two perpendicular linearly

polarized components with arbitrary amplitudes and phase constants, and alternatively in terms of two oppositely circularly polarized components with arbitrary amplitudes and phase constants. In Sec. 4.3 we discuss how one can produce a desired polarization state, either by control of the emission process, by elimination of an undesired component through selective absorption, or by enhancement of a desired component by selective reflection. In Sec. 4.4 we discuss double refraction in quarter- and half-wave plates and the analogous phenomena of optical activity and of relative phase shifts in metallic reflection. In Sec. 4.5 we discuss the relation between bandwidth $\Delta\nu$, coherence time $(\Delta\nu)^{-1}$, and polarization. Then we are able to say what we mean by "unpolarized" light or "partially polarized" light.

Chapter 5 - Interference and Diffraction. We begin (Sec. 5.2) with the interference pattern of two coherent slit (or point) sources. In Sec. 5.3 we consider the analogous interference pattern from two independent slit sources each having bandwidth $\Delta\nu$. We find that the interference pattern remains fixed for approximately the coherence time $(\Delta\nu)^{-1}$ and can be measured during that time by a sufficiently fast technique. In Sec. 5.4 we apply the concepts of Secs. 5.2 and 5.3 to learn how large a "point" (or how wide a "line") source can be and still behave like a true point source when used to illuminate a double slit. An easy home experiment (number 17, Chap. 5) illustrates the "loss of coherence" when the point gets too large. Another easy home experiment (number 18) illustrates coherence with Lloyd's mirror. In Sec. 5.5 we examine "beams" of traveling waves produced in various ways. We learn that such a beam (like a searchlight beam) can never be exactly "unidirectional" but has an angular spread in its direction of approximately λ/D, where λ is the wavelength and D the spatial width of the beam. In Sec. 5.6 we consider the interference pattern of many parallel slits and the diffractional pattern of a single wide slit in an opaque screen. We also discuss the way in which an opaque screen works, emphasizing the point of view that all electromagnetic radiation has its origin in sources of oscillating charges. We are led to Huygens' construction in such a way that we recognize that it is an approximation that should not work very well in the immediate vicinity of the edges of the slit. We are also led to an understand-

ing of the diffraction from an opaque object in a light beam. In Sec. 5.7 we study "geometrical optics" from a wave point of view. We first show that the law of specular reflection is the result of constructive interference, as is Snell's law of refraction. Then we study an ellipsoidal mirror with a point source at one focus, which gives an image at the other focus that is not a point but is of radius about $\frac{1}{4}\lambda$. Then we stretch one focus of the ellipsoid to infinity to obtain a parabolic mirror and see that a point source at the focus gives a beam that is almost unidirectional(but not quite, unless the aperture of the mirror is infinite). The angular width satisfies the (general) relation $\Delta\theta \approx \lambda/D$. Next we examine thin prisms and thin lenses. We find that the general condition for any kind of lens (focussing light or, perhaps, charged particles) to focus a parallel beam to a point (neglecting now the diffraction size of the "point") is that the angular deviation of a ray towards the axis be linearly proportional to the lateral displacement of the ray from the axis. In problems 22, 23, and 24 the student is asked to design a telescope that will enable him to improve his mailing-tube spectrometer (Home Exp. 20) to the point of being able to resolve the two sodium D lines (Home Exp. 19).

Note: Students and instructors are warmly encouraged to send me suggestions. If they reach me before Christmas 1966, they will influence the final edition. If not, send them anyway. An easy and valuable technique is to ask the students for a one or two-page "book review" as a substitute for a problem set. Send the reviews (not summaries of them) to:

Professor Frank S. Crawford, Jr.
Department of Physics
University of California
Berkeley, California 94720.

CONTENTS

Preface to Volume III. iii

Chapter 1 - Modes.

1.1 Introduction. 1-1

1.2 Oscillations of a system with 1 degree of freedom. 1-10

1.3 Linearity and the Superposition Principle. 1-19

1.4 Oscillations of a system with two degrees of freedom. 1-22

1.5 Beats. 1-34

1.6 Systems with many degrees of freedom. Dispersion relations. 1-48

Problems. 1-70

Chapter 2 - Waves.

2.1 Introduction. 2-1

2.2 Harmonic traveling waves. 2-5

2.3 Superposition of two harmonic traveling waves. Group velocity. 2-15

2.4 Superposition of many harmonic waves. Pulses and wave packets. 2-23

2.5 Wave equations. 2-40

Problems. 2-59

Chapter 3 - Emission and absorption of waves: Impedance.

3.1 Introduction. 3-1

3.2 Emission and absorption of traveling waves on a continuous string. 3-8

3.3 Sound waves. 3-22

*3.4 Current and voltage waves on a transmission line. 3-47

*3.5 Radiation by an oscillating sheet of charge. 3-65

3.6 The general case of electromagnetic plane waves in vacuum. 3-72

*3.7 Reflection at a reactive load. 3-80

Problems and Home Experiments. 3-98

Chapter 4 - Polarization.

4.1 Introduction. 4-1

4.2 Description of polarization. 4-6

4.3 Production of polarized transverse waves. 4-15

4.4 Double refraction. 4-29

4.5 Bandwidth, coherence time, and polarization. 4-40

Problems and Home Experiments. 4-51

Chapter 5 - Interference and Diffraction

5.1 Introduction. 5-1

5.2 Interference between two coherent point sources. 5-2

5.3 Interference between two independent sources. 5-13

5.4 How large can a "point" light source be? 5-20

5.5 Angular width of a "beam" of traveling waves -- approximate treatment. 5-24

5.6 Multiple-slit interference pattern and single-slit diffraction pattern. 5-31

5.7 Geometrical optics. 5-45

Problems and Home Experiments. 5-62

Advanced Topics.

1. "Microscopic" examples of weakly-coupled identical oscillators. AT-1

2. Dispersion relation for de Broglie waves. AT-5

3. Phase and group velocities for de Broglie waves. AT-8

4. Wave equations for de Broglie waves. AT-10

5. Three-dimensional classical wave equation with spherical symmetry. AT-11

6. Reflection of de Broglie waves at discontinuities in potential energy. AT-15

7. Solution of driven oscillator problem by Green's function method. AT-19

Appendix. A-1

Chapter 1 - Modes

Section		Page
1.1	Introduction.	1
1.2	Oscillations of a system with 1 degree of freedom.	10
1.3	Linearity and the Superposition Principle.	19
1.4	Oscillations of a system with two degrees of freedom.	22
1.5	Beats.	34
1.6	Systems with many degrees of freedom. Dispersion relations.	48
	Problems.	70

1.1 Introduction.

The world is full of things that move. Their motions can be broadly categorized into two classes, according to whether the thing that is moving stays near one place or travels from one place to another. Examples of the former are an oscillating pendulum; a vibrating violin string; water sloshing back and forth in a cup; electrons vibrating (or whatever they do) in atoms; light bouncing back and forth between the mirrors of a laser. Parallel examples of traveling motion are a sliding hockey puck; a pulse traveling down a long stretched rope plucked at one end; ocean waves rolling towards the beach; the electron beam of a TV tube; a ray of light emitted at a star and detected at your eye. Sometimes the same phenomenon exhibits one or the other class of motion (i.e., standing still on the average, or traveling) depending on your point of view: the ocean waves travel toward the beach, but the water (and the duck sitting on the surface) goes up and down (and also forwards and backwards) without traveling. The displacement pulse travels down the rope, but the material of the rope vibrates without traveling. In the case of the light traveling from the star to your eye, it's not so simple to say what it is that doesn't travel. A hundred years ago we might have said that the "ether" stands still and vibrates like jelly while the light beam travels. Now all we say is that the inertial frame of the observer stands still, and the light travels through the empty vacuum of the inertial frame (and also through transparent substances like air and water).

We shall begin by studying things that stay in one vicinity and oscillate or vibrate about an average position. In Sec. 1.2 we shall consider some very simple systems each of which has only a single "moving part," for example, a pendulum oscillating in a plane, a mass on a spring, and an LC

circuit. (See Fig. 1.) Such a system, whose configuration at any time can

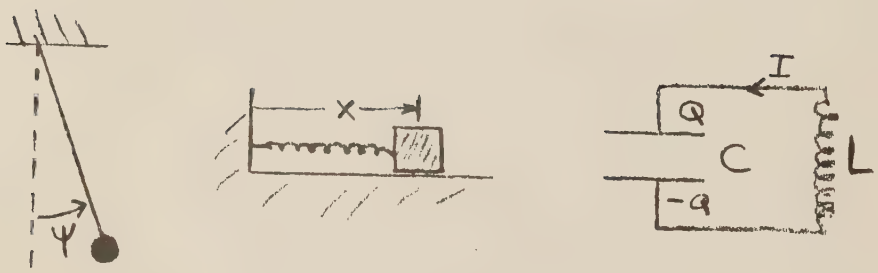


Fig. 1-1. Systems with one degree of freedom. (The pendulum is constrained to swing in a plane.)

be completely specified by giving a single quantity, is said to have one degree of freedom -- loosely speaking, one moving part. For example, the swinging pendulum can be described by the angle that the string makes with the vertical, and the LC circuit by the charge on the condenser. (A pendulum free to swing in any direction, like a bob on a string, has not one but two degrees of freedom; it takes two coordinates to specify the position of the bob. The pendulum on a grandfather clock is constrained to swing in a plane, and thus has only one degree of freedom.)

For all of these systems having one degree of freedom we shall find that the displacement of the "moving part" from its equilibrium value has the same simple time dependence (called harmonic oscillation)

$$\psi(t) = A \cos \omega t \ , \qquad (1)$$

where ψ may represent the displacement of the oscillating mass from its equilibrium position, or it may represent the charge on the condenser for the oscillating LC circuit. More precisely, we shall find the time dependence to be given by Eq. (1) provided the moving part does not move too far from its equilibrium position. [For large-angle swings of a pendulum, Eq. (1) is a poor approximation to the motion; for large extensions of a real spring the return force is not proportional to the extension, and the motion is not given by Eq. (1); a large enough charge on a condenser will cause it to "break down" by sparking between the plates , and the charge will not satisfy Eq. (1)].

The oscillatory behavior represented by Eq. (1) is always the result of two intrinsic properties of the physical system which have opposing tendencies. One of these properties is a "return force" which tries to return ψ to zero by imparting a suitable "velocity" $d\psi/dt$ to the moving part. The greater ψ is, the stronger the return force is. For example, for the oscillating LC circuit, the "return force" is due to the repulsive force between the electrons, which makes the electrons prefer not to crowd onto one of the condenser plates, but rather to distribute themselves equally on each plate, giving "zero charge." (The "charge" on a condenser always means the charge <u>transferred</u> from one plate to the other.) The second property is "inertia," which "opposes" any change in $d\psi/dt$. For example, for the oscillating LC circuit the inertia is due to the inductance L , which opposes any change in the current $d\psi/dt$ (letting ψ stand for the charge on the condenser). The interplay of these two tendencies, return force and inertia, leads to oscillatory behavior. For example, if we start with ψ positive and $d\psi/dt$ zero, the return force gives an acceleration which induces a negative velocity. By the time ψ returns to zero the negative velocity is maximum. The return force becomes zero at $\psi = 0$,

but the negative velocity now induces a negative displacement. Then the return force becomes positive, but it must now overcome the inertia of the negative velocity. Finally, the velocity $d\psi/dt$ becomes zero, but by that time the displacement is large and negative, and the process reverses. This cycle goes on and on: The return force tries to restore ψ to zero, in so doing it induces a velocity; the inertia preserves the velocity and causes ψ to "overshoot." Thus the system oscillates.

Though the interplay of return force and inertia always leads to oscillatory behavior, it does not always lead to Eq. (1). Only if the "acceleration" $d^2\psi/dt^2$ is linearly proportional to the negative of the displacement ψ (and not proportional, for example, to $-\psi^3$) do we get Eq. (1). We examine this and other consequences of linearity in Sec. 1.3.

After we have examined several examples of systems with one degree of freedom, we shall study systems with two degrees of freedom in Sec. 1.4. Examples are a double pendulum (one pendulum attached to the ceiling, the second attached to the bob of the first);two pendulums coupled by a spring; a string with two beads; two coupled LC circuits; or a pendulum free to swing in any direction. (See Fig. 2.) It takes two variables to describe the configuration of such a system, say ψ_a and ψ_b. For example, in the case

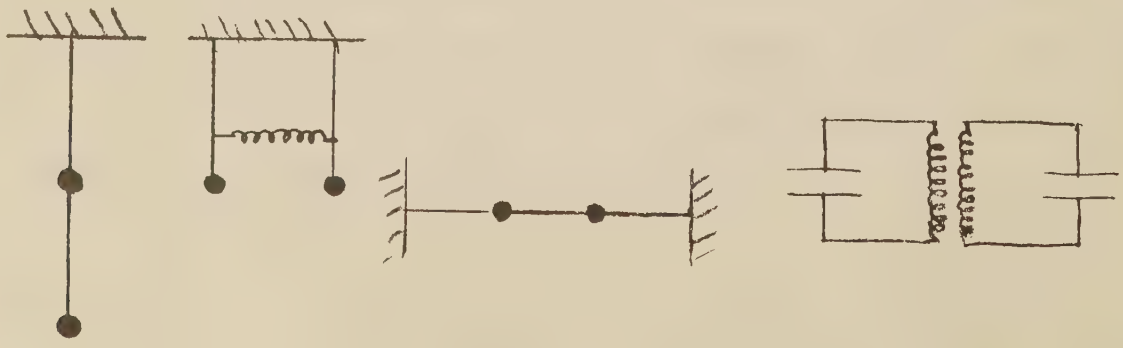


Fig. 1-2. Systems with two degrees of freedom. (The masses are constrained to remain in the plane of the figure.)

of a pendulum free to swing in any direction, the "moving parts" ψ_a and ψ_b would be the positions of the pendulum in the two perpendicular horizontal directions; in the case of coupled pendulums, the moving parts ψ_a and ψ_b would be the positions of each pendulum; in the case of two coupled LC circuits the "moving parts" ψ_a and ψ_b would be the charges on the two condensers.

The general motion of a system with two degrees of freedom can have a very complicated appearance; no part moves with simple harmonic motion. However, underlying the apparent complexity there are two (for two degrees of freedom) remarkably simple motions, both going on simultaneously. These simple motions (described below) are called "normal modes," or simply "modes." By suitable starting conditions (suitable values of ψ_a , ψ_b , $d\psi_a/dt$, and $d\psi_b/dt$) we can get the system to oscillate in only one mode or the other.

When only one mode is present, each moving part undergoes simple harmonic motion, as given by Eq. (1). All parts oscillate with the same frequency. All parts pass through their equilibrium positions (where ψ is zero) simultaneously. Thus, in a single mode, one never has, for example, $\psi_a(t) = A \cos \omega t$ and $\psi_b(t) = B \sin \omega t$ or $\psi_a(t) = A \cos \omega_1 t$ and $\psi_b(t) = B \cos \omega_2 t (\omega_1 \neq \omega_2)$. Instead one has, for one mode (which we call mode 1)

$$\psi_a(t) = A_1 \cos \omega_1 t \, ,$$

$$\psi_b(t) = B_1 \cos \omega_1 t \, . \qquad (2)$$

Similarly for mode 2 the two moving parts a and b have

$$\psi_a(t) = A_2 \cos \omega_2 t \ ,$$
$$\psi_b(t) = B_2 \cos \omega_2 t \ . \qquad (3)$$

Each mode has its own characteristic frequency, ω_1 for mode 1, ω_2 for mode 2. Each mode also has its own characteristic ratio of amplitudes of the motion of the moving parts, A_1/B_1 for mode 1, and A_2/B_2 for mode 2. Note that the ratio $\psi_a(t)/\psi_b(t)$ is constant, independent of time, in a mode. It is given by the appropriate ratio A_1/B_1 or A_2/B_2, which can be either positive or negative.

The most general motion of the system is simply a superposition with both modes oscillating at once:

$$\psi_a(t) = A_1 \cos \omega_1 t + A_2 \cos (\omega_2 t + \phi) \ ,$$
$$\psi_b(t) = B_1 \cos \omega_1 t + B_2 \cos (\omega_2 t + \phi) \ , \qquad (4)$$

where the phase constant ϕ allows for the fact that mode 2 need not be "turned on at the same time" as mode 1 (in Eqs. (2) and (3) we arbitrarily "set the clock" so that the displacements were maximum at time $t = 0$; we cannot do that for both modes in a general starting condition).

After studying several examples of systems with two degrees of freedom, we can generalize to any number of degrees of freedom f . For a system with f degrees of freedom, there are always exactly f modes, each having its own frequency ω and its own amplitude ratios A:B:C:D: ... etc.,corresponding to the degrees of freedom a,b,c,d, ... etc. In each mode all moving parts

go through their equilibrium positions simultaneously (i.e., all have the same phase constant). The moving parts a,b,c, etc., will have motion in a given mode described by the natural extension of Eq. (2) or Eq. (3). We can always pick out one of the moving parts, say a , and take its amplitude in mode 1, A_1 , as "the amplitude of mode 1," because once A_1 is fixed then B_1, C_1, D_1, etc., are fixed by the ratios characteristic of mode 1. The amplitudes of the other modes can be characterized similarly by the amplitude of one moving part. The most general motion of the system is a superposition of all its modes, with the amplitude and phase constant of each mode set in the initial conditions. The appearance of the vibrating system, for a typical situation, is very complicated, simply because the eye and brain cannot contemplate several things at once without confusion. It is not easy to look at the complete motion and "see" each mode separately when many are present.

In Sec. 1.5 we make a short detour to study beats. We start by considering beats between the two normal modes of a system of 2 degrees of freedom and then go on to study some other varieties of beats.

In Sec. 1.6 we return to our study of modes. We consider transverse vibrations of a beaded string with many beads. We are mainly interested to see what happens as we increase the number of degrees of freedom, until we have a huge number of them. If a system contains a very large number of moving parts, and if these parts are distributed within a limited region of space, then the average distance between neighboring moving parts becomes very small. As an approximation one may wish to think of the number of parts as becoming infinite and the distance between neighboring parts as going to zero. One then says that the system behaves as if it were "continuous."

Implicit in this point of view is the assumption that the motion of near neighbors is nearly the same. This assumption allows us to describe the displacement of all of the moving parts in a small neighborhood of a point x, y, z, with a single quantity $\psi(x,y,z,t)$. Then the "displacement" $\psi(x,y,z,t)$ is a continuous function of position, x, y, z, and of time t . It replaces the displacements $\psi_a(t)$, $\psi_b(t)$, etc., of the individual parts. We then say we are dealing with waves.

Nature does not actually exhibit any continuous material systems. For example, air is not continuous, since it consists of discrete air molecules separated (at room temperature and at 1 atmosphere pressure) by about 3×10^{-7} cm on the average, which is about 30 molecular diameters. Nevertheless, to a very good approximation air may be considered continuous in any situation where a given small neighborhood contains a huge number of air molecules all undergoing the same average motion. If for a suitable "huge number" we take 10^6 , then the size of the "small neighborhood" is that of a cube of edge about 3×10^{-5} cm. For a sound wave, the average motion of molecules is about the same over distances small compared to the wavelength. The highest audible notes have wavelength about 1.5 cm, which is large compared to our "small neighborhood." Thus when we study sound waves in air we shall treat the air as continuous. Similarly we shall study idealized violin strings using the approximation of continuity.

The modes of a continuous system are called standing waves, or normal modes, or simply modes. According to the above discussion a truly continuous system has an infinite number of independent moving parts, occupying a finite space. There are therefore an infinite number of degrees of freedom, and hence an infinite number of modes. This is indeed the case for the electromagnetic field in vacuum. It is not true for a real material system.

One liter of air does not contain an infinite number of moving parts, but only 2.7×10^{22} molecules, each of which has three degrees of freedom (for motion along x, y, and z directions). Thus a bottle containing 1 liter of air does not have an infinite number of possible vibrational modes of the air, but only 8×10^{22} at most. Anyone who has practiced blowing a bottle or a flute knows that it is not easy to excite more than the first few modes. (We always number the modes by calling the one with the lowest frequency number 1, the next higher number 2, etc.) In practice we are usually only concerned with the first few (or few dozen or few thousand) modes, and the lowest modes always behave as if the system were continuous.

In our study of the modes of a beaded string in Sec. 1.6 we shall encounter our first example of a dispersion relation, which is just a relation between frequency and wave length. Perhaps you are now raising your eyebrows and saying to yourself, "I learned in high school that the relation between frequency and wavelength for waves is very simple: their product is the wave velocity" (for traveling waves). That is true. But the velocity need not be a constant, independent of wavelength. Therefore if you have $\lambda\nu = v(\lambda)$, i.e., v is some function of wavelength, then you don't know the relation between wavelength and frequency until you know this functional dependence. We shall find the dispersion relation for the beaded string and will learn that in the limit of a continuous string (an infinite number of beads in a finite length) the traveling-wave velocity does become independent of wavelength. It becomes the constant $(T/\rho)^{\frac{1}{2}}$, where T is the string tension and ρ is the mass per unit length of the string.

1.2 Oscillations of a system with 1 degree of freedom.

The first three examples in this section are for review. (They are given in greater detail in Volume 1, Chap. 7.) We shall consider a simple pendulum, a mass sliding on a frictionless table under the influence of a spring, and an LC oscillating circuit. For simplicity we shall in all cases neglect friction (or resistance in the case of the LC circuit). We shall in every case find that for small enough displacements $\psi(t)$ satisfies the differential equation

$$(\text{inertial constant}) \cdot \frac{d^2\psi(t)}{dt^2} = -(\text{force constant}) \cdot \psi(t) \qquad (5)$$

which has the general solution

$$\psi(t) = A \cos(\omega t + \phi) , \qquad (6)$$

where A is the amplitude (usually taken to be positive), $\omega = 2\pi\nu$ is the angular frequency in radians per second, ν is the frequency in cycles per second (abbreviated cps); ν is the inverse of the period T, which is given in seconds per cycle:

$$\nu = 1/T . \qquad (7)$$

The angular frequency ω is related to the physical properties of the system by

$$\omega^2 = \frac{\text{force constant}}{\text{inertial constant}} . \qquad (8)$$

The phase constant ϕ corresponds to the choice of the origin of time. We can always "reset the clock" so that ϕ becomes zero, e.g., by using t' instead of t, with $\omega t' = \omega t + \phi$, so that t' differs from t by the constant ϕ/ω. Therefore we usually set $\phi = 0$. Now let us consider the physical examples:

Pendulum. A simple pendulum consists of a massless string or rod of length ℓ attached at the top to a rigid support and at the bottom to a "point" bob of mass M. (See Fig. 3.) Let ψ denote the angle (in radians) that the string makes with the vertical. (The pendulum swings in a plane; its configuration is given by ψ alone.)

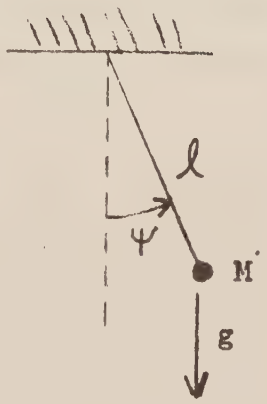


Fig. 1-3. Simple pendulum

The component of acceleration of the bob tangential to the circular arc of its path is $\ell d^2\psi/dt^2$. The return force is the tangential component of the weight Mg, namely $-Mg \sin \psi$. Thus Newton's second law (mass times acceleration equals force) gives

$$\frac{M\ell d^2\psi}{dt^2} = -Mg \sin \psi(t) \quad . \tag{9}$$

Equation (9) is not of the form of Eq. (5), because $\sin \psi$ is not ψ. But we can use the Taylor's series expansion [Appendix, Eq. (4)]

$$\sin \psi = \psi - \frac{\psi^3}{3!} + \frac{\psi^5}{5!} - \cdots \tag{10}$$

where the ellipsis (···) denotes the rest of the infinite series. We see that for sufficiently small ψ (in radians, remember) we can neglect all terms in Eq. (10) except the first one, ψ . You may ask, "How small is 'sufficiently small'?" That question has no universal answer -- it depends on how accurately you can determine the function $\psi(t)$ in the experiment you have in mind (this is physics, remember -- nothing is perfectly measurable), and how much you care. For ψ = 0.10 rad (5.7 deg), sin ψ is 0.09983; in some problems this is a poor approximation. For ψ = 1.0 rad (57.3 deg), sin ψ is 0.841; in some problems this is an adequate approximation.

If we retain only the first term in Eq. (10), then Eq. (9) takes on the same form as Eq. (5). Thus the solution is of the form Eq. (6), with ω^2 given by Eq. (8) as

$$\omega^2 = g/\ell \quad . \tag{11}$$

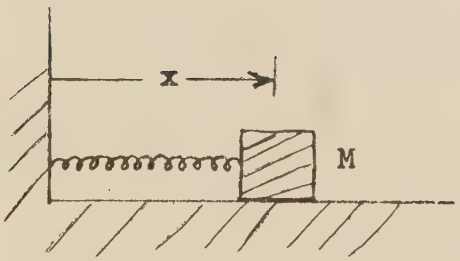


Fig. 1-4. Mass sliding on frictionless surface, with force exerted by a spring.

Mass and spring. Suppose that x is the length of the spring, and x_o is the length when the spring is relaxed. (See Fig. 4.) Suppose the potential energy stored in the spring is given by the potential function $V(x)$ (not necessarily the potential function of an idealized spring, $\frac{1}{2} Cx^2$) . We can expand $V(x)$ in a Taylor's series at $x = x_o$ [Appendix, Eq. (3)]. Thus

$$V(x) = V(x_o) + (x-x_o)(dV/dx)_o + \frac{(x-x_o)^2}{2!}(d^2V/dx^2)_o + \frac{(x-x_o)^3}{3!}(d^3V/dx^3)_o + \ldots \quad (12)$$

where the subscript zeros on the derivatives mean that they are evaluated at $x = x_o$. Since $x = x_o$ is an equilibrium point (the return force is zero when the spring is relaxed), we have $(dV/dx)_o = 0$. The return force on the mass is given by [Volume 1, p. 138, Eq. (14)]

$$F_x = - dV/dx \ ,$$

evaluated at x, not x_o. We let ψ denote the displacement $x - x_o$ from the equilibrium point. Then, by differentiating Eq. (12) with respect to x we can write Newton's second law as

$$\frac{Md^2\psi}{dt^2} = F_x = - (d^2V/dx^2)_o\psi - \tfrac{1}{2}(d^3V/dx^3)_o\psi^2 - \ldots \ . \quad (13)$$

For sufficiently small ψ we may neglect all but the first term in the expansion. Then we get Eq. (5). The solution is again given by Eq. (6), with ω^2 given by Eq. (8) as

$$\omega^2 = \frac{(d^2V/dx^2)_o}{M} \ . \quad (14)$$

Note that in both the preceding examples a non-linear restoring force is approximated for small displacements by the linear term in the corresponding Taylor's expansion. This is a common situation in diverse physical systems, where the "restoring force" may not necessarily be merely a mechanical force, but may be of a much more general nature.

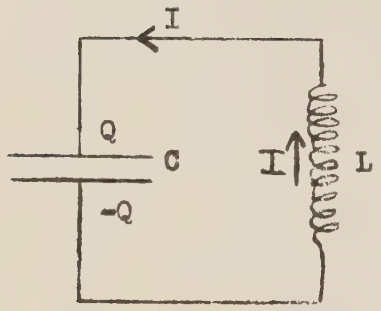


Fig. 1-5. Resonant LC circuit. The sign conventions for Q and I are indicated.

LC Circuit. (For a more complete discussion see Volume 2, Chap. 8.) Consider the series LC circuit of Fig. 5. The charge displaced from the bottom to the top plate of the condenser is Q. At equilibrium Q is zero. When Q is positive (for example), there is a "return force" proportional to the electromotive force (potential difference) V across the condenser. We can write (Volume 2, Sec. 3.5)

$$V = -C^{-1}Q \qquad (15)$$

where the minus sign tells us that V tends to return Q to zero. If we think of V as a return force, then C^{-1} is like a spring constant. (We should keep in mind that this is only an analogy. The dimensions of Q are those of charge, not distance, so that V does not have the dimensions of force, nor C^{-1} the dimensions of a spring constant.)

The current I is positive (by our convention) when it has the direction of the arrow in Fig. 5. Thus

$$I = dQ/dt \ . \qquad (16)$$

The inductance L has the "inertial" property that it "tries" to preserve dQ/dt. Thus LdQ/dt is somewhat analogous to Mdx/dt in the example of

the mass and spring. The analogue of "mass times acceleration equals force" is

$$\frac{Ld^2Q}{dt^2} = -C^{-1}Q \quad . \tag{17}$$

This has the form of Eq. (5) with ψ the charge displacement Q, L the inertial constant, and C the force constant. The angular frequency is given by

$$\omega^2 = \frac{C^{-1}}{L} \quad . \tag{18}$$

The charge and current oscillate [as in Eq. (6)] according to

$$Q = Q_o \cos \omega t \tag{19}$$

$$I = \frac{dQ}{dt} = -\omega Q_o \sin \omega t \quad . \tag{20}$$

Plasma oscillations. A neutral plasma consists of a gas of neutral molecules with some of the molecules ionized. Every singly-ionized molecule consists of one positive ion which has set free one (negative) electron. An important example of a plasma is the earth's ionosphere, a layer of air (actually several layers with somewhat differing properties) that contains many ionized air molecules (N_2 and O_2 molecules). The ionization of an air molecule usually occurs by means of absorption of an ultra-violet light quantum emitted by the sun. The ion and free-electron density is greatest at about 200 to 400 km above the earth's surface. Higher than that, the electron (and

ion) density decreases because the density of neutral air molecules available to be ionized decreases. Lower than that, the electron density decreases because the ultra-violet radiation has already been mostly absorbed. (We would soon die of sunburn without the protective air layer above us.)

Since the plasma is neutral (on the average), it does not act as a source of electrostatic field. However one region of the plasma may have at any one instant a slight excess of charge, with a corresponding deficit in some other neighboring region. This creates an electric field in the plasma. Under the influence of the electric field the ions are accelerated in one direction (i.e., along the field) and the electrons in the other. The charges move in directions which tend to cancel the charge excesses and deficits which created the electric field. Thus we have a "return force." By the time the charge excess has been cancelled to zero and the corresponding electric field reduced to zero the ions and electrons have acquired velocities. Their inertia causes them to "overshoot," and we get a new charge excess and deficit with the opposite sign to the original one. We have here the typical situation that leads to oscillations.

If we are interested only in the net transport of charge from one region to another, we can forget about the positive ions and consider the entire motion of charge to be due to the motion of the electrons. This is because the acceleration of an electron is larger than that of a singly-charged ion by the ratio of their respective masses (about 3×10^4), since the electric force is the same.

Let us study a simplified situation where the plasma is confined between the plates of a condenser. We neglect motion of the ions. At any instant there may be an excess of charge Q on one plate and a corresponding deficit on the other plate. This gives a uniform electric field in the plasma (Volume 2, Sec. 3.5)

Fig. 1-6. Plasma oscillations in a condenser.

$$E_x = -4\pi Q/A \tag{21}$$

where A is the area of the condenser plate, and the minus sign signifies that E_x tends to return Q to zero. There is no other source of electric field. (The plasma between the plates is neutral, because for every electron that moves out of a given region, another moves in.) A single electron has mass m and charge $q = -e = -4.803 \times 10^{-10}$ esu . Newton's second law gives

$$\frac{md^2x}{dt^2} = qE_x \tag{22}$$

for every electron in the plasma. (We are neglecting other forces on the electrons which arise from collisions among the electrons and ions; these forces average to zero and give no net motion of charge.) Now suppose that there are N free electrons per cubic centimeter and that each electron is

displaced from its average (equilibrium) position by a distance x. Then the net charge deposited on one plate (and removed from the other) is given by

$$Q = NqAx \quad . \tag{23}$$

Differentiating Eq. (23) twice with respect to time and inserting Eqs. (22) and (21) again yields an equation like Eq. (5), namely,

$$\frac{d^2Q}{dt^2} = -\frac{4\pi\, Nq^2}{m} Q \quad . \tag{24}$$

This has a solution of the form of Eq. (6) with angular frequency given by Eq. (8) as

$$\omega^2 = \frac{4\pi\, Ne^2}{m} \equiv \omega_p^2 \quad . \tag{25}$$

The quantity ω_p is called the plasma oscillation frequency. (It is sometimes customary to call ω merely "frequency" instead of the more complete "angular frequency." If someone says "the frequency is so many radians per second," you know they mean ω and not ν, which is given in cps.)

The free-electron density N in the earth's ionosphere varies with height and with time. After sunset the recombination of ions and electrons to form neutral molecules continues, whereas the forming of new ions ceases. Therefore the electron density decreases at night. Typical daytime plasma oscillation frequencies are

$$\nu_p = 10 \text{ to } 30 \text{ Mc} \quad \text{(Megacycles per second)} \quad , \tag{26}$$

which corresponds to $N \approx 10^6$ to 10^7 free electrons per cm^3.

1.3 Linearity and the Superposition Principle.

In Sec. 1.2 we solved for the oscillations of the pendulum and of the mass and spring only for the case where the displacement ψ was small enough so that we could assume the return force to be proportional to $-\psi$, with (for example) no dependence on ψ^2, ψ^3, etc. A differential equation that contains no higher than the first power of ψ, $d\psi/dt$, $d^2\psi/dt^2$, etc., is said to be linear in ψ and its time derivatives. If, in addition, no terms independent of ψ occur, the equation is said to be homogeneous as well. If any higher powers of ψ or its derivatives occur in the equation, the equation is said to be non-linear. For example, Eq. (9) is non-linear, as we can see from the expansion of $\sin\psi$ given by Eq. (10). So is Eq. (13) non-linear. Only when we neglect the higher powers of ψ do we obtain a linear equation.

Non-linear equations are generally very difficult to solve. (The non-linear pendulum equation, Eq. (9), is solved exactly in Volume 1, p. 225 ff.) Fortunately there are many interesting physical situations for which linear equations give a very good approximation. We shall deal mostly with linear equations.

Linear homogeneous differential equations have the following very interesting and important property: The sum of any two solutions is itself a solution. This is easily proven as follows: Suppose $\psi_1(t)$ and $\psi_2(t)$ are solutions of the linear homogeneous equation

$$\frac{d^2\psi(t)}{dt^2} = -C\,\psi(t) \quad , \tag{27}$$

where C is a constant. Thus by hypothesis

$$\frac{d^2\psi_1(t)}{dt^2} = - C \psi_1(t)$$

$$\frac{d^2\psi_2(t)}{dt^2} = - C \psi_2(t) \quad .$$

Adding these two equations, we obtain

$$\frac{d^2[\psi_1(t) + \psi_2(t)]}{dt^2} = - C[\psi_1(t) + \psi_2(t)] \tag{28}$$

which has the same form as Eq. (27) with

$$\psi(t) = \psi_1(t) + \psi_2(t) \quad . \tag{29}$$

Thus $\psi_1 + \psi_2$ is a solution if ψ_1 and ψ_2 are solutions. The sum of two solutions of a linear equation is often called a "superposition" of the two solutions. The property that a superposition of solutions is itself a solution is unique to homogeneous linear equations. Oscillations that obey such equations are said to obey the superposition principle.

As an example of a differential equation to which the superposition principle does not apply, consider the exact non-linear differential equation of the pendulum,

$$\frac{d^2\psi(t)}{dt^2} = - C \sin \psi(t) \quad . \tag{30}$$

Suppose $\psi_1(t)$ and $\psi_2(t)$ are two solutions of Eq. (30). Is $\psi_1 + \psi_2$ a solution? The answer is no, as is easily seen: given

$$\frac{d^2\psi_1(t)}{dt^2} = - C \sin \psi_1(t)$$

$$\frac{d^2\psi_2(t)}{dt^2} = - C \sin \psi_2(t) \quad .$$

Adding these two equations we obtain

$$\frac{d^2[\psi_1(t) + \psi_2(t)]}{dt^2} = - C[\sin \psi_1(t) + \sin \psi_2(t)] \quad . \tag{31}$$

If it were true that

$$\sin \psi_1(t) + \sin \psi_2(t) = \sin [\psi_1(t) + \psi_2(t)]$$

then we could write Eq. (31) in the form of Eq. (30), with $\psi(t) = \psi_1(t) + \psi_2(t)$. However this does not hold.

Linear inhomogeneous equations (i.e., equations containing terms independent of ψ) also give rise to a superposition principle, though of a slightly different sort. Many physical situations occur that are analogous to that of a driven harmonic oscillator, which satisfies

$$\frac{Md^2\psi(t)}{dt^2} = - C\psi(t) + F(t) \quad , \tag{32}$$

where $F(t)$ is an "external" driving force that is independent of $\psi(t)$.

The corresponding superposition principle is as follows: Suppose a driving force $F_1(t)$ produces an oscillation $\psi_1(t)$, [when F_1 is the only driving force], and suppose another driving force $F_2(t)$ produces an oscillation $\psi_2(t)$ [when F_2 is present by itself]. Then, if both driving forces are present simultaneously [so that the total driving force is the superposition $F_1(t) + F_2(t)$] , the corresponding oscillation [i.e., corresponding solution of Eq. (32)] is given by the superposition $\psi(t) = \psi_1(t) + \psi_2(t)$. We leave it to you to show that this is true for the linear inhomogeneous Eq. (32), and not true for an equation non-linear in $\psi(t)$.

1.4 Oscillations of a system with two degrees of freedom.

There are many fascinating examples in nature of systems having two degrees of freedom. The most beautiful examples involve molecules and elementary particles (the neutral K mesons especially). Their study requires quantum mechanics. We shall only consider a few simple mechanical examples here.

Our main goal is to give you a feeling for normal modes, so that, for example, you can guess the properties of the modes of simple systems without working out the mathematics. We shall first consider a system composed of a stretched string with two beads (fixed on the string) which can undergo transverse vibrations. We shall study this system in detail, and then give further examples where we only guess the modes.

Consider the beaded string of Fig. 7. Both beads have mass M . All three strings have equilibrium length ℓ_o and tension T_o . When the beads are not at their equilibrium positions, ψ_a is the displacement of bead a , and ψ_b the displacement of b . The length of

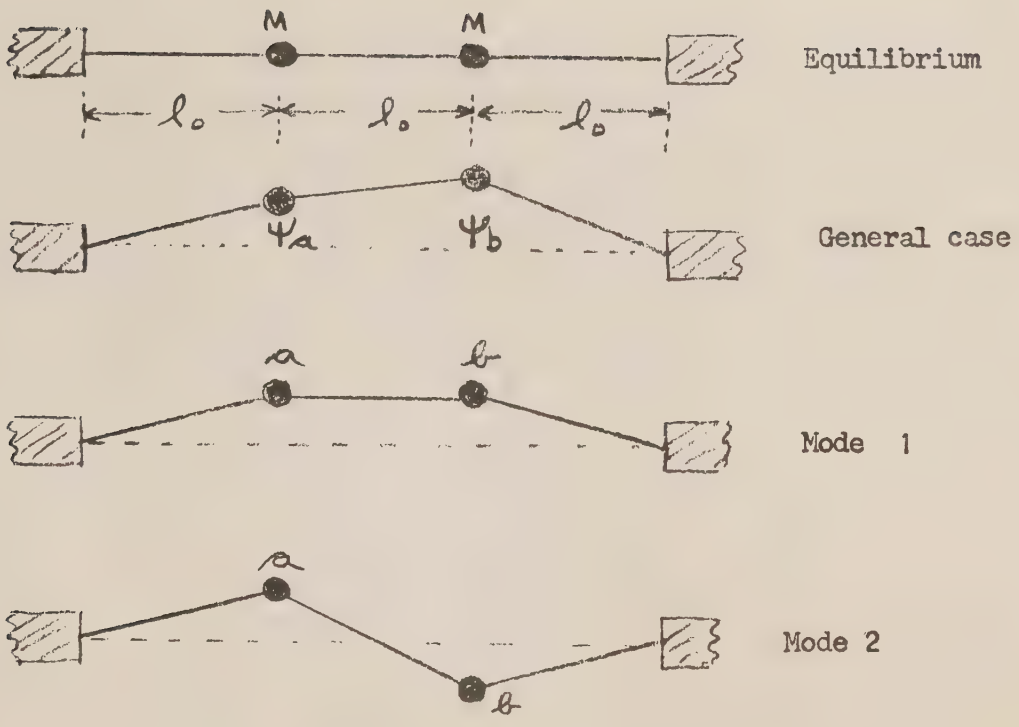


Fig. 1-7. Vibration of beaded string with two beads.

the first string becomes $(\ell_o^2 + \psi_a^2)^{\frac{1}{2}}$, with similar expressions for the lengths of the other two strings. The tensions also increase.

We wish to deal only with linear equations. It turns out that if we consider only small values of displacement, such that

$$|\psi_a/\ell_o| \ll 1 , \quad \text{and} \quad |\psi_b/\ell_o| \ll 1 , \tag{33}$$

then we can regard the lengths of the strings as constant, ℓ_o, and the tensions as constant, T_o. (See problem 5, part (c). That problem also contains, in part (b), a very special case which gives a linear equation even for large-amplitude vibrations.) The transverse force exerted by any string is the tension T_o times the sine of the angle the string makes with

the equilibrium string axis. For small oscillations, as defined by Eq. (33), the sine of the angle equals the angle (in radians) and also equals the tangent of the angle. [See Appendix Eqs. (4) and (19).] Thus the first string exerts force $-T_o(\psi_a/\ell_o)$ on mass a. The second string exerts force $+T_o(\psi_b - \psi_a)/\ell_o$ on mass a, and $-T_o(\psi_b - \psi_a)/\ell_o$ on mass b. The third string exerts force $-T_o\psi_b/\ell_o$ on mass b. Newton's second law gives, therefore,

$$\frac{Md^2\psi_a}{dt^2} = -\frac{T_o}{\ell_o}\psi_a + \frac{T_o}{\ell_o}(\psi_b - \psi_a) \tag{34}$$

$$\frac{Md^2\psi_b}{dt^2} = -\frac{T_o}{\ell_o}(\psi_b - \psi_a) - \frac{T_o}{\ell_o}\psi_b \ . \tag{35}$$

These are formidable-looking equations. It is clear that we have no right to expect either ψ_a or ψ_b to undergo harmonic oscillation, because for ψ_a, for example, the force is not a "return force," i.e., it is not proportional to $-\psi_a$, but includes a contribution proportional to ψ_b. The only chance for the return force on ψ_a to be proportional to ψ_a is for ψ_b to be just a constant times ψ_a. If that is the case then the force on ψ_b is also proportional to ψ_b, and thus ψ_b can also undergo harmonic oscillation. It is clear that if ψ_a and ψ_b are to maintain a constant ratio, they must oscillate with the same frequency. Thus, if we are to have harmonic oscillation, the right-hand side of Eq. (34) must equal $-C\psi_a$, and the right-hand side of Eq. (35) must equal $-C\psi_b$, with the same constant C. The frequency will then be given by

$$\omega^2 = C/M \quad . \tag{36}$$

Let us try to find C . Setting the right side of Eq. (34) equal to $-C\psi_a$, and that of Eq. (35) equal to $-C\psi_b$, and collecting coefficients of ψ_a and ψ_b , we obtain

$$(C - 2T_o/\ell_o)\ \psi_a + \frac{T_o}{\ell_o}\psi_b = 0 \quad . \tag{37}$$

$$\frac{T_o}{\ell_o}\psi_a + (C - 2T_o/\ell_o)\ \psi_b = 0 \quad . \tag{38}$$

Solving Eq. (37) for ψ_b/ψ_a , and then solving Eq.(38) for ψ_b/ψ_a, we obtain

$$\frac{\psi_b}{\psi_a} = -\frac{(C - 2T_o/\ell_o)}{T_o/\ell_o} = -\frac{T_o/\ell_o}{(C - 2T_o/\ell_o)} \quad . \tag{39}$$

The second equality of Eq. (39) gives

$$(C - 2T_o/\ell_o)^2 = (T_o/\ell_o)^2 \quad .$$

Taking the square roots we obtain two solutions C_1 and C_2 given by

$$C_1 = T_o/\ell_o \quad , \tag{40}$$

$$C_2 = 3T_o/\ell_o \quad . \tag{41}$$

Inserting these solutions into Eq. (36) gives the corresponding frequencies

$$\text{mode 1:}\quad \omega_1^2 = T_o/M\ell_o \ ,$$
$$\text{mode 2:}\quad \omega_2^2 = 3T_o/M\ell_o \ . \tag{42}$$

From Eq. (39) we find the corresponding amplitude ratios

$$\text{mode 1:}\quad \psi_b/\psi_a = 1 \ ,$$
$$\text{mode 2:}\quad \psi_b/\psi_a = -1 \ . \tag{43}$$

The two modes are shown in Fig. 7.

We have found the two modes without thinking, just by solving equations. Let us see how we can guess them (of course, knowing the answer helps). Our reasoning will be applicable to systems having any number of degrees of freedom, even though more complex systems are more difficult to guess. We try to guess a configuration such that the ratio of displacements ψ_a/ψ_b will remain constant in time. In other words the "shape" of the system remains constant, and there is, so to speak, just one overall oscillation. To maintain ψ_a/ψ_b constant, the "return" acceleration per unit displacement must be the same for all the beads. Thus a bead that has twice as great a displacement as another must also have twice as great an acceleration, so that it can maintain twice as great a velocity, so that it can cover twice as much space in any time interval. Now, the return acceleration per unit displacement is just the return force per unit mass per unit displacement. It is that which must be the same for all beads in a mode.

In the present example the beads have the same mass. Therefore we need

only require that the return force per unit displacement be the same for both beads. A little study of Fig. 5 shows that the two configurations shown for modes 1 and 2 satisfy that condition. The symmetry of the system (i.e., equal masses, equal string lengths, same conditions at each end -- namely tied to a support) helps very much in guessing the configurations of the modes.

Once we have guessed the configurations ("shapes") of the modes, we can find the frequencies without going through the algebra of Eqs. (37) through (43). Since in a given mode every particle (bead) has the same frequency, we can use any one bead. The return force per unit mass per unit displacement is ω^2, in a mode. This you can easily see; it is exactly what one has for any harmonic oscillation, e.g., for the mass on the spring; and it is Eq. (8). By inspection of Fig. 5, the transverse return force on bead a in mode 1 is $T_o \psi_a / \ell_o$ (using the small-angle approximation where the tangent, sine, and angle in radians are all equal). The return force per unit displacement per unit mass is just $T_o/\ell_o M$, which is just ω_1^2. This agrees with the algebraic result of Eq. (42). Similarly for mode 2, inspection of Fig. 5 shows us that, for bead a, the second string exerts twice the return force that the first string exerts, and in the same direction, thus giving a total of $1 + 2 = 3$ times $T_o/\ell_o M$ for ω_2^2. This also agrees with Eq.(42).

Now we are ready to look at the general solution, which is a superposition of modes 1 and 2 with arbitrary amplitude for each mode (which we can call the amplitude for bead a, since, given the mode, we know ψ_b/ψ_a) and arbitrary turn-on time (phase constant) for each mode. Thus, let us say that ψ_a is a superposition of a contribution $A_1 \cos(\omega_1 t + \phi_1)$ from mode

1 and a contribution $A_2 \cos(\omega_2 t + \phi_2)$ from mode 2. There is no more freedom -- ψ_b is completely specified by the initial conditions that correspond to the choices A_1, A_2, ϕ_1, and ϕ_2. Thus, ψ_b is a superposition of a contribution $A_1 \cos(\omega_1 t + \phi_1)$ from mode 1 (because $\psi_b = \psi_a$ in mode 1) plus a contribution $-A_2 \cos(\omega_2 t + \phi_2)$ from mode 2 (because $\psi_b = -\psi_a$ in mode 2). Putting this all together we have

$$\psi_a(t) = A_1 \cos(\omega_1 t + \phi_1) + A_2 \cos(\omega_2 t + \phi_2) \qquad (44a)$$

$$\psi_b(t) = A_1 \cos(\omega_1 t + \phi_1) - A_2 \cos(\omega_2 t + \phi_2) \; . \qquad (44b)$$

Note that the superposition $\psi_a + \psi_b = 2A_1 \cos(\omega_1 t + \phi_1)$ is an oscillation with the frequency of mode 1, and that the superposition $\psi_a - \psi_b = 2A_2 \cos(\omega_2 t + \phi_2)$ is an oscillation with the frequency of mode 2. We can call $\psi_a + \psi_b$ by the name ψ_1, the oscillation of mode 1, and $\psi_a - \psi_b$ by the name ψ_2, the oscillation of mode 2.

How do we know that Eqs. (44) represent the most general solution of the problem? Is it possible that there is some ingenious vibration that can <u>not</u> be represented by Eqs. (44) with suitable choices of the four constants A_1, A_2, ϕ_1, and ϕ_2? The answer is no. We first point out that, given the initial displacements and transverse velocities of both beads at time $t = 0$, we can solve for the four constants. Conversely, given the four constants we can find the initial displacements and velocities. (See problem 6.) Thus any set of initial conditions can be represented by the four constants.

Your physical intuition may tell you that the initial positions and

velocities for a system of any number of moving parts will determine the motion from there on, and that is correct. Given the initial velocities, we can calculate the changes in the displacements over the first tiny time interval Δt . (For this we can assume the velocities are essentially constant.) Thus we can find the new displacements at time Δt . Also, given the initial displacements, we know from Newton's second law (Eqs. (34) and (35) in the above example) the initial accelerations. Using these we can find the changes in velocities over the interval Δt . (For this we can assume the accelerations constant.) Thus we can find the new velocities at time Δt . Now we have a complete new set of "initial conditions" at time Δt and can proceed just as we did at time zero. We can go on in this way "forever." Thus specification of the initial positions and velocities completely determines the state of the system for all time. Also, since any set of initial conditions for our system can be represented by the four constants in our solutions, Eqs. (44), these solutions can be adjusted to behave in any way possible for the system. Q.E.D.

*General case. Why are there exactly two modes for two degrees of freedom? More generally, why are there exactly f modes for f degrees of freedom? An isolated system of f moving parts (i.e., f degrees of freedom) will obey a set of f linear homogeneous second order differential equations -- namely, Newton's second law for each moving part (at least, one can write a linear approximation for each moving part using the first term of the Taylor's series). The most general motion is determined by specifying the displacement and velocity for each moving part at $t = 0$. Thus a set of $2f$ constants describes any possible motion of the system. Now, each mode has two constants which are at our disposal, namely, the "amount" and "turn-on

time" (amplitude and phase constant). Thus it takes at least f modes, giving $2f$ constants in the solution, to adapt the solution to any possible set of initial conditions of the system (and thus to describe any possible motions). No more than f modes can be found, since f modes can be adapted to any possible set of initial conditions.

How do you know that as many as f modes exist for a system of f degrees of freedom? To answer this question, we examine the method by which we explicitly found the two modes for our example with $f = 2$, and then generalize the method to arbitrary f. The two linear homogeneous equations in the two unknowns ψ_a and ψ_b, Eqs. (40) and (41), have a unique solution for the ratio ψ_a/ψ_b if and only if the determinant of the coefficients vanishes, i.e., if and only if we have

$$\begin{vmatrix} (C - 2T_o/\ell_o) & T_o/\ell_o \\ T_o/\ell_o & (C - 2T_o/\ell_o) \end{vmatrix} = 0 \ . \qquad (45)$$

In fact, Eq. (45) is identical with the second equality of Eq. (39). Equation (45) is a quadratic equation in C. It has two solutions C_1 and C_2, yielding two modes. Similarly, if we had three equations analogous to the two of Eqs. (34) and (35), i.e., for three beads, then we would have had three homogeneous linear equations corresponding to Eqs. (37) and (38). These would have a unique solution (for the ratios $\psi_a : \psi_b : \psi_c$) provided the corresponding determinant vanishes. The "determinant equals zero" condition leads to a cubic in C, with three solutions C_1, C_2, and C_3, and three modes.

That is the way it goes: For f degrees of freedom there are f

simultaneous linear homogeneous second order differential equations. Assuming a mode in which each moving part is oscillating with frequency ω implies that $Md^2\psi_a/dt^2$ is $-C\psi_a$, $Md^2\psi_b$ is $-C\psi_b$, etc. Putting these conditions into the differential equations gives us ordinary equations to solve -- f simultaneous linear homogeneous equations in the f unknowns ψ_a, ψ_b, ψ_c, ... According to the theory of linear equations there is a unique solution for the ratios $\psi_a : \psi_b : \psi_c : \ldots$ if and only if the determinant of the coefficients vanishes. The determinant is a polynomial of order f in the unknown quantity C; i.e., the determinant has the form

$$a_f C^f + a_{f-1} C^{f-1} + \ldots + a_2 C^2 + a_1 C + a_o = 0 \quad , \tag{46}$$

where the constants a_f, a_{f-1}, etc., depend on the physical parameters of the problem (masses, spring constants, coupling constants, etc.). Equation (46) has f solutions for C, which we call C_1, C_2, ... C_f. Each such solution for C corresponds to a different mode; e.g., substitution of C_1 back into the set of linear homogeneous equations gives the amplitude ratios $A_1 : B_1 : C_1 : \ldots$ for mode 1. Substitution of C_1 into Eq. (36) gives ω_1^2. Similarly the solution C_2 gives $A_2 : B_2 : C_2 : \ldots$ and gives ω_2^2. Thus we have proven that all f modes always exist by showing you explicitly how you can always find them.

The f normal modes of a linear system (i.e., one obeying the superposition principle) with f degrees of freedom are said to form a "complete set of functions" to describe the motion of the system. They are not the only complete set. One can construct other complete sets of functions by

linear combinations (superpositions) of normal modes. (Such a superposition is not a mode, since it contains oscillations of different frequencies.) One can in this way construct f new functions which also form a complete set. (In so doing one must make use of all of the modes and must form "independent" linear combinations.) There are an infinite number of ways to do this and a large number of books on the subject (with titles like "Linear Transformations").

In solving for the modes of a system, there are many different techniques. One is to solve the "determinant equals zero" equation for the frequencies of the modes, and then solve for the amplitude ratios. (In various branches of "linear" physics this equation goes by different names; two common names are "the secular equation," and "the characteristic equation." The determinant is often called the secular determinant. The frequencies of the modes are called by various names: "characteristic frequencies," and "eigenvalues," or "eigenfrequencies" are common. Everyone has his own jargon.) When you have a simple system which has only two or three or so degrees of freedom and which has some symmetries to help guide your intuition (like equal masses, equal string lengths, etc.), it is usually fastest and most rewarding to guess the configurations $\psi_a : \psi_b : \ldots$ and then find the frequencies by simple considerations. If your intuition breaks down you can always go back to algebra.

Here are two more examples.

Coupled pendulums. See Fig. 8. Each pendulum has length L and mass M, and the two are coupled by a spring of spring constant C at distance L' from the support. The pendulums hang vertically at equilibrium (i.e., the pendulum spacing equals the relaxed length of the spring). We shall let you

show that one has

$$\text{Mode 1:}\quad \psi_b = \psi_a\ ;\ \omega_1^2 = \frac{g}{L} \tag{47}$$

$$\text{Mode 2:}\quad \psi_b = -\psi_a\ ;\ \omega_2^2 = \frac{g}{L} + \frac{2C}{M}\frac{L'}{L} \tag{48}$$

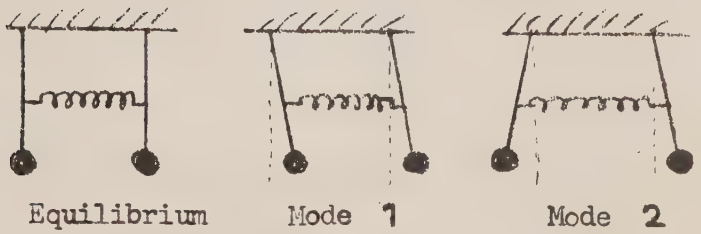


Fig. 1-8. Coupled pendulums.

Coupled Oscillators. See Fig. 9. Here it is easy to guess the currents and charges for the modes, but we need the differential equations to find the frequencies. We shall not pursue that. (See problem 7.)

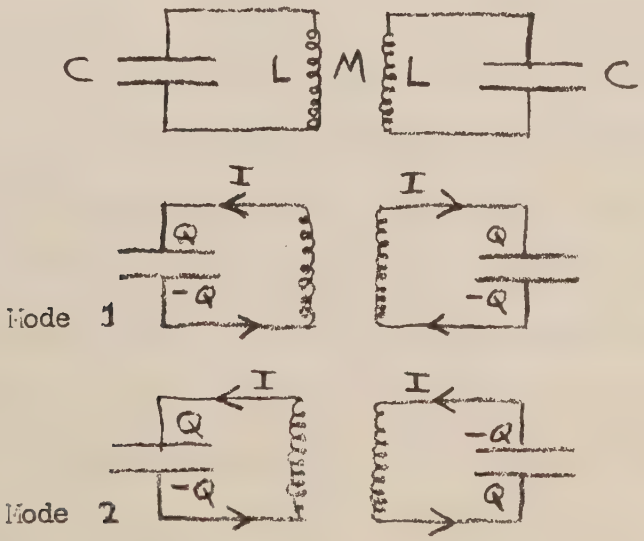


Fig. 1-9. Coupled oscillators. The magnetic flux in one coil produced by current in the other provides the coupling. It is represented by the mutual inductance M . There are many other ways of coupling the oscillators besides that shown here.

1.5 Beats.

In this section we depart from our orderly development of considering systems with 1 , 2 , ... degrees of freedom. You may, without loss of continuity, skip to Sec. 1.6, postponing the study of Sec. 1.5 until later.

Phenomena involving beats (to be explained in this section) occur in many branches of physics. We discuss beats at this time because one of their most interesting manifestations occurs in systems having two degrees of freedom. We shall start by describing beats that occur between the two normal modes of two weakly-coupled identical oscillators, leading to the interesting phenomenon of energy transfer from one oscillator to the other and back again. Later we shall discuss beats from systems of a different physical nature. But we shall find that the mathematics is the same. Lastly we shall discuss "combination tones," (which include beats) produced by non-linear (non-Hooke's law) response of a driven oscillator.

Weakly-coupled identical oscillators. If the spring that couples the identical pendulums in Fig. 8 is very weak (i.e., if C is small), or if it is attached very near to the ceiling (i.e., if L'/L is small), then the two pendulums are nearly independent. They are "weakly coupled." Then the frequencies ω_1 and ω_2 of the modes are nearly the same [see Eqs. (47) and (48)], and $|\omega_1 - \omega_2|$ is small compared with either ω_1 or ω_2. If you have such a situation and start at $t = 0$ with just one of the pendulums vibrating, say pendulum a, the other being at rest, a very fascinating process unfolds. Gradually the oscillation amplitude of pendulum a decreases and that of b increases, until eventually a is resting and b is oscillating with the amplitude and energy that a started out

with. (We are neglecting frictional forces.) The vibration energy has been transferred completely from one pendulum to the other. By the symmetry of the system we see that the process will continue. The vibration energy will slowly "flow" back and forth between a and b . One complete round trip for the energy from a to b and back to a is called one "beat." The beat period is the time for the round trip, and is the inverse of the beat frequency. The angular beat frequency (2π times the beat frequency) is given by

$$\omega_{beat} = |\omega_1 - \omega_2| \quad . \tag{49}$$

The phenomenon of beats is illustrated in Fig. 10 with ω_{beat} about 1/10 ω_1 . In this figure ψ_1 is $\psi_a + \psi_b$, the oscillation of mode 1, and ψ_2 is $\psi_a - \psi_b$, the oscillation of mode 2. Thus the superposition $\psi_1 + \psi_2$ is twice the oscillation ψ_a of pendulum a . (The superposition $\psi_1 - \psi_2$ is twice the oscillation ψ_b of pendulum b).

Here is one way to think of this energy transfer process. We think of pendulum b (initially at rest) as a driven pendulum. The driving force is provided by pendulum a , pushing on b through the intermediary of the coupling spring. The driving frequency is essentially equal to the natural frequency of the driven pendulum (both pendulums swing at almost the frequency of a free pendulum). Therefore b is driven at its resonant frequency, and its amplitude at first builds up linearly with time. If the driving pendulum a had a huge mass compared to the driven pendulum b , and hence a huge store of energy, it would be hardly affected by the loss of energy to b . We would have a conventional oscillator driven at resonance. Eventually the

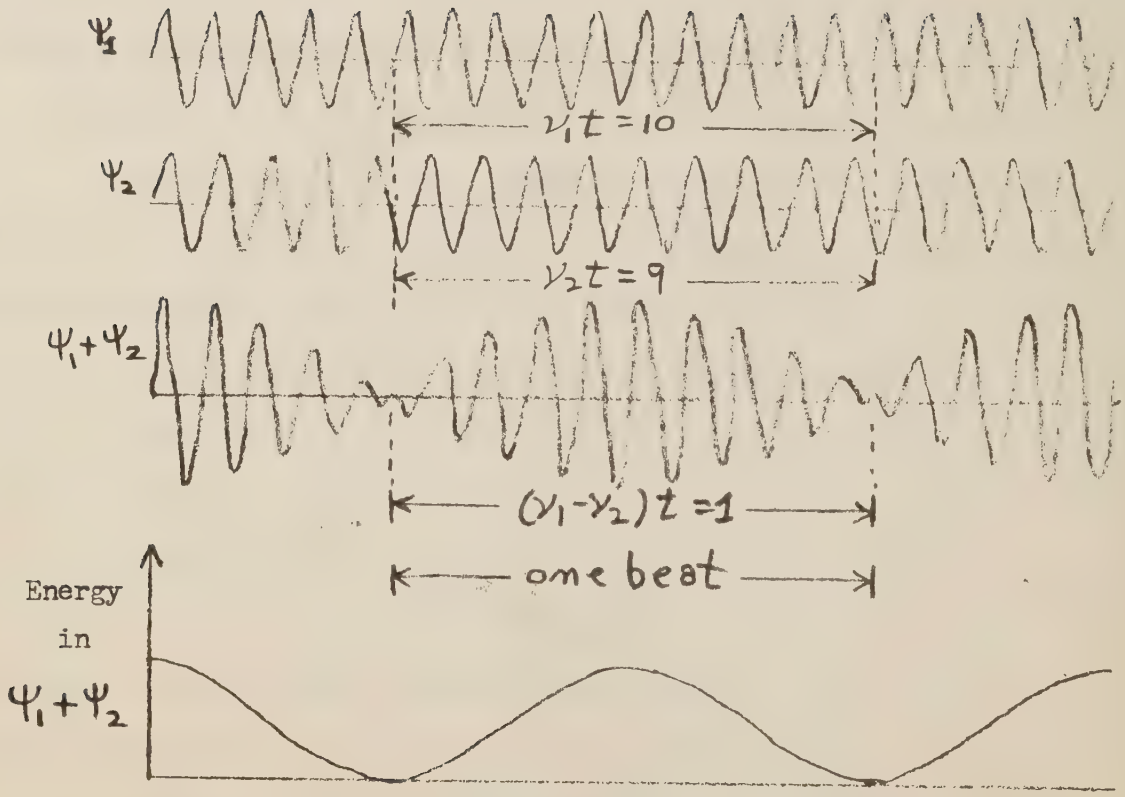


Fig. 1-10. Beats. Beats result from the superposition $\psi_1 + \psi_2$ of two oscillations ψ_1 and ψ_2 that have nearly the same frequency. If the amplitudes of ψ_1 and ψ_2 are equal, the amplitude of $\psi_1 + \psi_2$ becomes zero once per beat. (If they are not equal, one still has beats, but the minimum amplitude of the superposition is not zero.) The superposition $\psi_1 - \psi_2$ (not shown) also exhibits beats, but displaced by $\frac{1}{2}$ of a beat cycle from those of $\psi_1 + \psi_2$. The sum of the energy in $\psi_1 + \psi_2$ plus that in $\psi_1 - \psi_2$ is constant.

amplitude of b would become so large that the (previously neglected) friction would take as much energy away from b on each swing as is fed in by a. But for our case, pendulum a does <u>not</u> have a huge mass compared

to b , so the store of energy of a dies away until the roles become reversed.

This point of view can be confusing. For example, when a has lost half its energy, b and a have the same energy. How does the energy then "know" which way to flow? How does it know which (of a and b) is the driven oscillator and which the driving force? The answer is that the relative phases of a and b are still the same no matter what fraction of its energy a has lost. Pendulum a is still trying to "hurry up" b , and b is trying to "slow down" a , until the energy transfer is complete. (See problem 10.)

Here is another way to think of the energy transfer process. The two normal modes have the same "shapes" as the modes for the two-bead string, as one can see from Eqs. (47) and (48). Consequently the solutions will be of the form of Eqs. (44) with the appropriate frequencies from Eqs. (47) and (48). Their frequencies differ slightly, by an amount $\Delta\omega = \omega_2 - \omega_1$. The amplitudes and phase constants must be set so that at $t = 0$ pendulum b is at rest at its equilibrium position. According to Eqs. (44) we can satisfy the initial conditions by letting both phase constants be zero and letting A_1 and A_2 be equal. Thus the starting conditions give

$$\psi_a(t) = A \cos \omega_1 t + A \cos \omega_2 t \qquad (50a)$$

$$\psi_b(t) = A \cos \omega_1 t - A \cos \omega_2 t \quad . \qquad (50b)$$

For small times (small compared with a beat period), since $\omega_1 \approx \omega_2$, Eqs. (50) give

$$\psi_a(t) \approx 2A \cos \omega_1 t \tag{51a}$$

$$\psi_b(t) \approx 0 \ . \tag{51b}$$

Instead of writing $\psi_a(t)$ and $\psi_b(t)$ as the superpositions given by Eq. (50) we can write them in a form that expresses the fact that there is a fast frequency of oscillation and a slow beat frequency at which the amplitude varies. Let us define the average frequency ω_{av} by

$$\omega_{av} = \tfrac{1}{2}(\omega_1 + \omega_2) \ . \tag{52}$$

Then in Eqs. (50) we can write

$$\omega_1 = \omega_{av} - \tfrac{1}{2}(\omega_2 - \omega_1) \tag{53}$$

$$\omega_2 = \omega_{av} + \tfrac{1}{2}(\omega_2 - \omega_1) \ , \tag{54}$$

where we take ω_2 to be the larger of ω_1 and ω_2. Then, making use of the identities

$$\cos (a - b) + \cos (a + b) = 2 \cos a \cos b \tag{55}$$

$$\cos (a - b) - \cos (a + b) = 2 \sin a \sin b \tag{56}$$

with $a = \omega_{av} t$ and $b = \tfrac{1}{2}(\omega_2 - \omega_1)t$, we obtain from Eqs. (50) the result

$$\psi_a(t) = [2A \cos \tfrac{1}{2}(\omega_2 - \omega_1)t] \cos \omega_{av} t \tag{57}$$

$$\psi_b(t) = [2A \sin \tfrac{1}{2}(\omega_2 - \omega_1)t] \sin \omega_{av} t \ . \tag{58}$$

Equations (57) and (58) have an interesting form. The factors in square brackets can be thought of as slowly varying amplitudes. Thus

$$\psi_a(t) = A(t) \cos \omega_{av} t \tag{59}$$

$$\psi_b(t) = B(t) \sin \omega_{av} t \ , \tag{60}$$

with $A(t)$ and $B(t)$ given by the bracketed quantities in Eqs. (57) and (58):

$$A(t) = 2A \cos \tfrac{1}{2}(\omega_2 - \omega_1)t \tag{61}$$

$$B(t) = 2A \sin \tfrac{1}{2}(\omega_2 - \omega_1)t \ . \tag{62}$$

From Eqs. (59) and (60) we see that a and b are always 90 degrees out of phase on the fast oscillations at frequency ω_{av}. From Eqs. (61) and (62) we see that the slowly varying amplitudes $A(t)$ and $B(t)$ are 90 degrees out of phase at the slow frequency $\tfrac{1}{2}(\omega_2 - \omega_1)$.

You may wonder why the frequency of $A(t)$ and $B(t)$ is $\tfrac{1}{2}(\omega_2 - \omega_1)$ instead of $(\omega_2 - \omega_1)$. The answer is that the beat frequency $\omega_2 - \omega_1$ gives the frequency for the <u>energy</u> to make a round trip. The oscillation energy of pendulum a is large whenever $|A(t)|$ is large, whether or not $A(t)$ is positive or negative. The magnitude $|A(t)|$ is large when $\tfrac{1}{2}(\omega_2 - \omega_1)t = 0 \ , \ \pi \ , \ 2\pi \ , \ 3\pi$, etc., i.e., for $(\omega_2 - \omega_1)t = 0 \ , \ 2\pi \ , \ 4\pi$, etc. We see that $\omega_2 - \omega_1$ is the beat frequency and not $\tfrac{1}{2}(\omega_2 - \omega_1)$.

Let us find an expression for the energy (kinetic plus potential) stored in each of the pendulums as a function of time. The sum of the energy in the two pendulums is constant. (We neglect the energy in the coupling spring; you may show that for $|\omega_1 - \omega_2| \ll \omega_{av}$ the energy in the spring is tiny compared with that in the two pendulums.) The energy of one pendulum, for example pendulum a , varies slowly with time. If the coupling-spring constant were zero there would be no beats, and the energy of each pendulum would be constant, because $A(t)$ and $B(t)$ would be constant. To calculate

the energy of, say, pendulum a at a given time we treat $A(t)$ as if it were constant. Thus to calculate $d\psi_a/dt$ we write [starting with Eq. (59)]

$$\frac{d\psi_a}{dt} = \frac{d[A(t)\cos \omega_{av}t]}{dt} \approx A(t) \frac{d[\cos \omega_{av}t]}{dt} = -\omega_{av}A(t)\sin \omega_{av}t \ . \quad (63)$$

The velocity of pendulum a is $Ld\psi_a/dt$, where L is the pendulum length and ψ_a is the displacement angle. The kinetic energy is therefore

$$E_{kin}(a) = \tfrac{1}{2}ML^2(d\psi_a/dt)^2 = \tfrac{1}{2}ML^2\omega_{av}{}^2A^2(t)\sin^2\omega_{av}t \ . \quad (64)$$

The potential energy is $MgL\,[1 - \cos \psi_a]$, or, using the expansion [see Appendix, Eq. (5)]

$$\cos x = 1 - \tfrac{1}{2}x^2 + \ldots \approx 1 - \tfrac{1}{2}x^2 \quad \text{for} \quad x^2 \ll 1 \ ,$$

$$E_{pot}(a) = \tfrac{1}{2}MgL\psi_a{}^2 = \tfrac{1}{2}MgLA^2(t)\cos^2\omega_{av}t \ . \quad (65)$$

The average angular frequency ω_{av} is approximately equal to the frequency of the free pendulums -- the frequency they would have if there were no coupling spring. Thus $\omega_{av}{}^2$ is approximately g/L [Eq. (11)], i.e., g is $L\omega_{av}{}^2$. Substituting this into Eq. (65) we see that the potential energy is similar in form to the kinetic energy, except that the kinetic and potential energies are 90 deg out of phase. That, of course, is because the velocity leads the displacement by 90 deg in phase. (Mnemonic: a given velocity induces, at a <u>later</u> time, a corresponding displacement. Thus the velocity "leads" the displacement.) Adding Eqs. (64) and (65) we then find

$$E_a = \tfrac{1}{2}ML^2\omega_{av}{}^2A^2(t)\,[\sin^2\omega_{av}t + \cos^2\omega_{av}t] \ , \quad (66)$$

Or, inserting the expression for $A(t)$ [Eq. (61)] into Eq. (66),

$$E_a = 2ML^2\omega_{av}{}^2A^2\cos^2 \tfrac{1}{2}(\omega_2 - \omega_1)t \quad . \tag{67a}$$

Similarly we find

$$E_b = 2ML^2\omega_{av}{}^2A^2\sin^2 \tfrac{1}{2}(\omega_2 - \omega_1)t \quad . \tag{67b}$$

We see that the sum of E_a and E_b is constant

$$E \equiv E_a + E_b = 2ML^2\omega_{av}{}^2A^2 \quad . \tag{68}$$

Finally we can rewrite Eqs. (67) in a form where the beat frequency, rather than half the beat frequency, appears explicitly. Using the identities

$$1 + \cos 2x = 2\cos^2 x \quad , \quad 1 - \cos 2x = 2\sin^2 x \tag{69}$$

with $x = \frac{1}{2}(\omega_2 - \omega_1)t$, we find

$$E_a = \tfrac{1}{2}E\,[1 + \cos(\omega_2 - \omega_1)t] \quad , \tag{70a}$$

$$E_b = \tfrac{1}{2}E\,[1 - \cos(\omega_2 - \omega_1)t] \quad . \tag{70b}$$

Thus the total energy E is constant, and flows back and forth between a and b with the beat frequency $|\omega_2 - \omega_1|$.

There are many ways by which you may demonstrate energy transfer between weakly coupled identical oscillators. Here are several you can do at home cheaply:

Pendulums. Use a washer or nut on a string. Couple the pendulums with a soda straw in place of the spring. Or hang a washer on a thread connecting the strings. Or use a rubber band to couple the strings.

Hacksaw blades. Clamp two hacksaw blades (about $0.25 apiece at a hardware store) with C-clamps to a table, with about four inches free to vibrate. One way to adjust them to the same frequency is to shorten the protruding part of one blade until it vibrates at a recognizable pitch, and then tune the other to sound the same pitch. Another way is to "strobe" each of them, using the light from a TV tube as a convenient stroboscope. (See problem 4.) When the two blades are in reasonably close tune, couple them with a rubber band.

Esoteric examples. In the study of microscopic systems -- molecules, elementary particles -- one encounters several beautiful examples of systems which are mathematically analogous to our mechanical example of two identical weakly-coupled pendulums. One needs quantum mechanics to understand these systems. The "stuff" that "flows" back and forth between the two degrees of freedom, in analogy to the energy transfer between two weakly coupled pendulums oscillating 90 deg out of phase, is not energy but probability. The energy is "quantized" -- it cannot "subdivide" to flow. Either one "moving part" or the other has all the energy. What "flows" is the probability to have the excitation energy. Two examples, the ammonia molecule (which is the "clockworks" of the ammonia clock) and the neutral K mesons, are discussed in Advanced Topic 1.

Beats of a different sort: two tuning forks. Obtain two tuning forks having the same nominal pitch. [The cheapest are about $1.25 at music stores. Both A440 (i.e., 440 cps) and C523.3 are readily available. For

the experiments we shall ask you to perform later C523.3 is more convenient, but A440 will do. Other forks are available from , e.g., Edmund Scientific Company, Barrington, N. J. 08007.] Strike one fork against the other at equal distances from the ends. Newton's third law (plus the symmetry of the striking condition) then insures equal vibration amplitudes for the forks, and also insures the same phase of the fast vibration at $t = 0$ for the two forks. Hold one near your ear. Then the other. Then both. The sound should be steady for either one, and should exhibit beats when you hold them both near one ear. (If they were not struck symmetrically, or are held at rather different distances from the ear, the minimum intensity will not be zero, because one can only get complete cancellation if the sound-wave amplitude produced at your ear by one fork is the same as that produced by the other.) In order to vary the beat frequency you may "load" one fork by wrapping a rubber band around one of its two prongs, adjusting the beat frequency by sliding the band nearer or farther from the end. The beats you hear are <u>not</u> the result of coupling two identical oscillators. (The two forks are practically not coupled at all. If one fork is struck and the other is not, and the forks are held near one another, the quiet one does <u>not</u> start to vibrate. There is no energy transfer between the forks.) Instead, these beats are simply the result of a superposition at your eardrum of the pressure variations (sound waves) produced by each fork. Suppose that one fork produces a pressure variation at your eardrum given by

$$p_1 = A \cos \omega_1 t \ , \tag{71}$$

and the other produces a pressure variation

$$p_2 = A \cos \omega_2 t \quad . \tag{72}$$

We have assumed the forks were struck symmetrically at $t = 0$, and are held about the same distance from your ear, so that we can neglect any phase constant and can also take the same amplitude A for both. The total pressure at your ear is the superposition of p_1 and p_2,

$$\begin{aligned} p = p_1 + p_2 &= A (\cos \omega_1 t + \cos \omega_2 t) \\ &= A \, \mathrm{Re} \, (e^{i\omega_1 t} + e^{i\omega_2 t}) \\ &= A \, \mathrm{Re} \, e^{i\frac{1}{2}(\omega_1 + \omega_2)t} \left[e^{i\frac{1}{2}(\omega_1 - \omega_2)t} + e^{-i\frac{1}{2}(\omega_1 - \omega_2)t} \right] \\ &= A \, \mathrm{Re} \, e^{i\omega_{av} t} \, 2 \cos \tfrac{1}{2}(\omega_1 - \omega_2)t \\ &= [2A \cos \tfrac{1}{2}(\omega_1 - \omega_2)t] \cos \omega_{av} t \quad . \end{aligned} \tag{73}$$

We see that the fast oscillation corresponding to ω_{av} is amplitude-modulated by a slowly-varying amplitude $A(t)$, given by

$$A(t) = 2A \cos \tfrac{1}{2}(\omega_1 - \omega_2)t \quad . \tag{74}$$

This $A(t)$ is identical in form with that given in Eq. (61).
In fact, the mathematics is identical with that which we used in discussing the beats between normal modes of weakly-coupled oscillators. But in the present case we don't have normal modes, just two forces (pressures)

which superpose to give constructive interference at some times (when $|A(t)| = 2A$) and completely destructive interference at other times (when $A(t) = 0$).

Figure 10 can correspond either to two weakly-coupled identical oscillators, as we examined above, or to the present case of superposition of sound waves from two tuning forks. In this case ψ_1 corresponds to the pressure variation due to one tuning fork, and ψ_2 is that due to the other. The superposition $\psi_1 + \psi_2$ is then the total pressure variation at your eardrum. Figure 10 can even correspond to beats between two light sources.*

If $|\nu_1 - \nu_2|$ is greater than about 10 or 20 cps the beats are too fast for your ear (plus brain) to notice. Instead, you hear two different notes, corresponding to ω_1 and ω_2. When that is the case there is no gain to our understanding in writing the superposition of p_1 and p_2 in the form of Eq. (73), even though Eq. (73) still holds exactly. Similarly, when we discussed energy transfer between weakly-coupled oscillators, we assumed we had $|\omega_1 - \omega_2| \ll \omega_{av}$. That was assumed so that the oscillators would undergo many fast oscillations during one beat and the beats would appear as slow variations in amplitude. In general the concept of beats is not very useful when the beat frequency is not small compared to the average frequency, because the motion becomes so complicated that one does not notice the beats.

---

* Photoelectric Mixing of Incoherent Light, by A. T. Forrester, R. A. Gudmundsen, and P. O. Johnson, Phys. Rev. 99, 1691 (1955). Their beautiful experiment used freely decaying mercury atoms with $\nu_1 - \nu_2 \approx 10^{10}$ cps and $\nu_{av} \approx 10^{15}$ cps

*Beats and "combination tones" produced by non-linear response. If the sound is not too loud your eardrum responds like a Hooke's law spring. The displacement is proportional to the pressure difference across the drum. For large enough displacements (i.e., loud enough sounds), a dependence on the square of the pressure difference becomes important. The response is then said to be non-linear. We can express this by writing

$$x(t) = Cp(t) + Dp^2(t) \quad . \tag{75}$$

For small enough pressure (difference) p, the quadratic term Dp^2 is negligible compared with the linear Hooke's law term Cp, and your brain does not notice it. Suppose that p is a superposition of two tones, one with (angular) frequency ω_1 and one with ω_2:

$$p = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t \quad . \tag{76}$$

If we substitute Eq. (76) in Eq. (75) and make use of the identity

$$\cos x \cos y = \tfrac{1}{2} [\cos (x + y) + \cos (x - y)] \tag{77}$$

we find

$$\begin{aligned} x(t) = {} & C [A_1 \cos \omega_1 t + A_2 \cos \omega_2 t)] \\ & + D [\tfrac{1}{2}A_1^2 \cos 2\omega_1 t + \tfrac{1}{2}A_2^2 \cos 2\omega_2 t + A_1 A_2 \cos (\omega_1+\omega_2)t \\ & \qquad + A_1 A_2 \cos (\omega_1-\omega_2)t] \\ & + D [\tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2^2] \quad . \end{aligned} \tag{78}$$

Now, your brain is capable of sorting out different frequency components of vibration of the eardrum. You can "listen" to one tone at a time by focusing your attention on it. This ability is usually practiced (in learning to sing, etc.) while listening to sounds of moderate loudness, where the non-linear (quadratic) response is negligible. But when it is not negligible, and when you get practiced at listening for faint tones in the presence of other loud tones, your brain can "recognize" in the vibration $x(t)$ given by Eq. (77) the tones corresponding to all of the frequencies ω_1 , ω_2 (both loud) and (faintly) $2\omega_1$, $2\omega_2$, $\omega_1 + \omega_2$, and $\omega_1 - \omega_2$. Frequencies $2\omega_1$ and $2\omega_2$ are called harmonics; $\omega_1 \pm \omega_2$ are called sum and difference tones.

What about the constant term $\frac{1}{2}D(A_1^2 + A_2^2)$? It has "zero frequency." It is just a steady displacement of the drum, similar to that which you get by swimming under water. Your brain does not interpret it as a sound. It just hurts.

It is possible to produce optical harmonics (and sum and difference frequencies) by making use of the non-linear response of a transparent substance. The cover of the Scientific American for July 1963 has a beautiful photograph showing a beam of red light of wavelength 6,940 angstrom units (1 angstrom = 10^{-8} cm) incident on a crystal. Emerging from the opposite side of the crystal is a beam of blue light having wavelength 3,470 angstrom (halving the wavelength is equivalent to doubling the frequency). (Actually, 3,470 angstrom corresponds to ultraviolet and is invisible to the eye. The color film used is sensitive to ultraviolet and the print comes out looking blue.) See also "The Interaction of Light with Light," by J. A. Giordmaine, Scientific American, April, 1964.

1.6 Systems with many degrees of freedom. Dispersion relations.

There are many interesting systems having a huge number of degrees of freedom that can be treated, to a good approximation, as "continuous." Some examples are a violin string and air at STP (standard temperature and pressure). In Chapter 3 we shall study vibrating violin strings and air columns, using the approximation that they are continuous. We shall then find that the standing waves of displacement of a continuous violin string (or of an air column) have the form

$$\psi(z,t) = A \sin kz \cos \omega t \quad . \tag{79}$$

(More generally we could write $A \cos (kz + \alpha) \cos (\omega t + \phi)$. But we set the clock so that ϕ is zero, and fix one end of the string at $z = 0$ so that α must be $+\pi/2$ or $-\pi/2$.) The quantity k is called the angular wave number. It has units of radians per unit length (rad/cm). (This is analogous to ω , which has units of rad/sec.) The quantity $k/2\pi$ is called the wave number. The wave number is the number of maxima (or minima) of the waveform per unit distance, analogous to the frequency, or number of maxima per unit time. Since the argument of a cosine function increases by 2π rad in one cycle, we can assign "units" of rad/cycle to the dimensionless number 2π. Then we see that the wave number $k/2\pi$ has units of (rad/cm)/(rad/cycle), which is cycles/cm. The inverse of the wave number is called the wavelength:

$$\text{wavelength} = \lambda = \frac{2\pi}{k} \quad . \tag{80}$$

The wavelength is the distance between successive maxima (or minima) of the wave form, i.e., the length of a complete cycle, analogous to the period T, or the time required for a complete cycle. The units of wavelength are the inverse of cycles/cm, i.e., cm/cycle (read this as "centimeters per cycle," not "centimeters over cycles").

When we study the continuous violin string and sound waves in air, we shall find that the ratio ω/k is constant. This ratio has units (rad/sec)/(rad/cm), which is cm/sec, i.e., the dimensions of velocity. When we study traveling waves, in Chap. 2, we shall find that ω/k is indeed the velocity of the traveling waves:

$$v = \frac{\omega}{k} \tag{81}$$

(Notice that ω/k equals λ/T, which is perhaps already familiar to you as an expression for wave velocity.) Thus, for a continuous violin string, and for ordinary sound waves, v is constant, independent of frequency. Velocity is also independent of frequency or wavelength for light in vacuum, where, as you know, we have $v = c = 3.0 \times 10^{10}$ cm/sec for all frequencies.

There are, however, many examples in nature where the wave velocity ω/k is not independent of frequency. The most familiar example is that of light traveling in a transparent medium like water, or glass, or air. The wave velocity depends on the frequency. The dependence is usually expressed (for light) in the form

$$v(\omega) = \frac{c}{n(\omega)} \tag{82}$$

where $n(\omega)$ is a function of ω called the index of refraction.

Equation (82) is an example of what is called a dispersion relation. Except in the case of light in a transparent medium, dispersion relations are not customarily written in the form of Eq. (82). Instead one simply says that the functional relation between ω and k, expressed either as $\omega(k)$ or as $k(\omega)$, is the dispersion relation. Because of the relation $v(k) = \omega(k)/k$ [or what amounts to the same thing, $v(\omega) = \omega/k(\omega)$], we need only know the functional relation between two of the three quantities v, ω, and k, to know the dispersion relation. A major part of the experimental and theoretical work in any branch of physics that involves systems with a huge number of degrees of freedom consists in trying to find (or guess) the dispersion relations.

In this section we shall study the normal modes of a beaded string. We shall find that, as the number of beads grows large and the bead spacing goes to zero, the modes take on the space and time dependence exhibited by Eq. (79). This tells us that the standing waves of a continuous violin string are nothing else but its normal modes. Further, in studying the beaded string we shall find our first example of a "non-trivial" dispersion relation, such that $v(k)$ is not independent of k. We shall also find that when the bead spacing becomes negligibly small compared to the wavelength, $v(k)$ becomes constant, independent of k (or what amounts to the same thing, independent of ω).

We begin by exhibiting in Fig. 11 the five modes of a string with five beads. We can easily imagine a smooth curve of the form $\sin(2\pi z/\lambda)$ which passes through all the beads (but does not follow the string exactly). We shall soon show that, for any beaded string having beads of equal mass

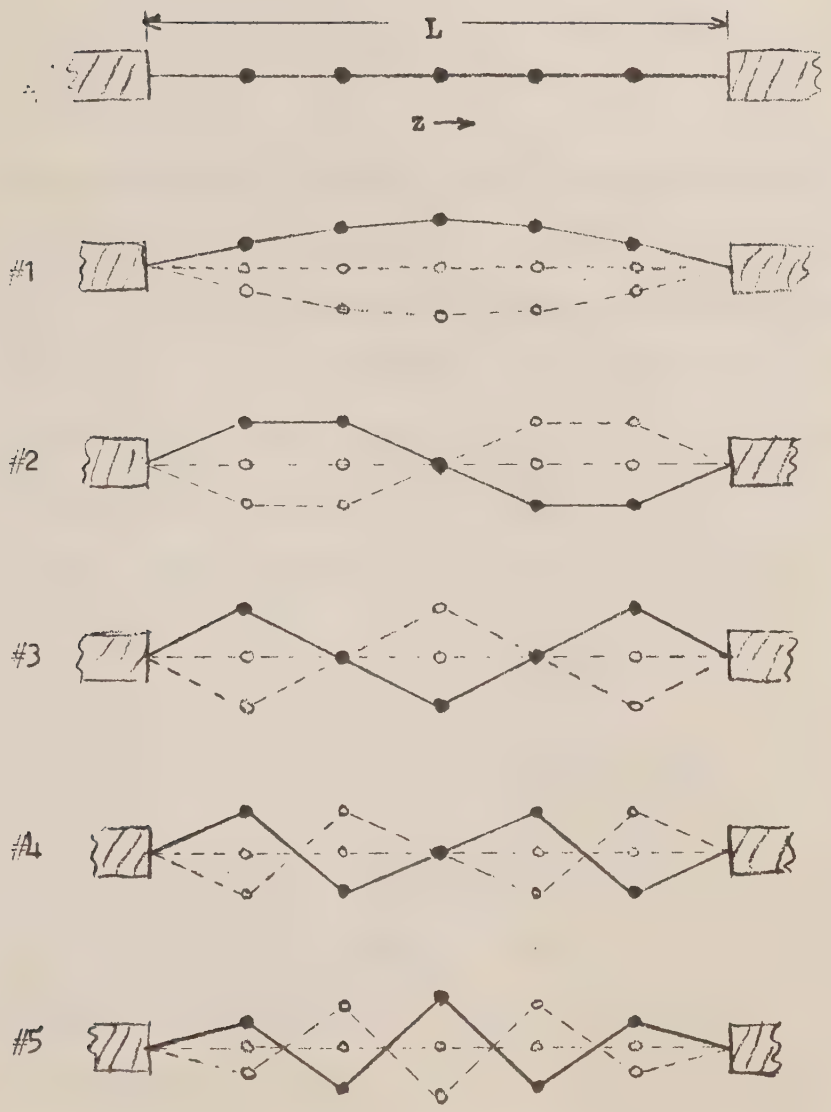


Fig. 1-11. The five modes of transverse vibration of a beaded string of five beads. The beads are constrained to vibrate in the plane of the figure. (For vibration in and out of the plane of the figure, there will be five more degrees of freedom, similar to the five shown.) For each mode, a smooth sine wave (not shown) can be passed through all of the beads for this or any analogous beaded string having beads of equal mass and string segments of equal length. For a string with many beads, the first five modes will look just like the smooth sine waves we can pass through the modes of the five-beaded string. For the many-beaded string the frequencies of the first few modes are in inverse ratio to the wavelength, i.e., are in the harmonic ratios 1:2:3:4:5 etc. This is not so for a modest number of beads, nor do the harmonic ratios prevail through the highest modes for a many-beaded string, except when "many" means "infinitely many."

and string segments of equal length, a curve of this form actually does pass through all the beads.

We have numbered the modes so that the total length L of the string is one-half the wavelength λ of the smooth curve $\sin(2\pi z/\lambda)$ for mode 1, two half wavelengths for mode 2, etc.; the mode number equals the number of half wavelengths. Thus λ depends on the mode. The lowest mode has the longest wavelength, $\lambda_1 = 2L$; the second mode has half that wavelength, $\lambda_2 = \frac{1}{2}\lambda_1$; etc:

$$\lambda_1 = 2L\ ,\ \lambda_2 = \frac{1}{2}\lambda_1\ ,\ \lambda_3 = \frac{1}{3}\lambda_1\ ,\ \lambda_4 = \frac{1}{4}\lambda_1\ ,\ \lambda_5 = \frac{1}{5}\lambda_1\ . \qquad (83)$$

The lowest mode vibrates the slowest: ω_1 is the smallest frequency, ω_2 is next, etc.: $\omega_1 < \omega_2 < \omega_3 < \omega_4 < \omega_5$. If the string had an infinite number of beads (keeping the length L) instead of five beads, we would have $\omega_2 = 2\,\omega_1$; $\omega_3 = 3\,\omega_1$; $\omega_4 = 4\,\omega_1$, etc. These simple integer ratios are called harmonics. For a modest number of beads one does not have harmonic ratios between the frequencies. For any number of beads one does have the integer ratios of Eq. (83) for the wavelengths of successive modes, but one has the corresponding inverse ratios for the frequencies only in the limit of a continuous string. To find the relation between the frequencies and the wavelengths, we must find the dispersion relation.

Now we shall solve exactly for the modes of a uniformly-beaded string. We shall not say what the total length of the string is, nor how many beads there are altogether, nor even whether the string is fixed at both ends (you can easily devise a constraint involving a frictionless ring sliding on a rod that will allow one end of the string to move freely in the transverse

direction, but will maintain the tension). We do demand that the return force on each bead be a linear restoring force. This means that the difference between displacements of neighboring beads must be small compared with the bead spacing (except for the special case that the strings are perfect Hooke's law springs having zero length before they are stretched to form the string segments -- see problem 5). It then turns out that we can neglect the increase in tension of a string segment when it is extended more than its equilibrium length. We thus take the tension to be a constant equal to the equilibrium value T.

It is the transverse component of T, $T \sin\theta$ (where θ is the angle the segment makes with the equilibrium direction), that provides the restoring force exerted by a string segment. We take $\theta \ll 1$ radian, and thus $\sin\theta = \theta = \tan\theta$. But $\tan\theta$ is the ratio given by the displacement increment (difference between displacements of each end of the segment) divided by the bead spacing a. Thus if we consider any three successive beads, which we call $n-1$, n, and $n+1$, the magnitude of the transverse tension exerted by the segment connecting $n-1$ and n is $T \tan\theta$, or $T\,|\psi_n - \psi_{n-1}|/a$. (See Fig. 12.) Thus the force exerted on bead n by the segment to its left in Fig. 12 is $-(T/a)[\psi_n - \psi_{n-1}]$; the force exerted by the segment to the right is $+(T/a)[\psi_{n+1} - \psi_n]$. The total force on bead n is therefore

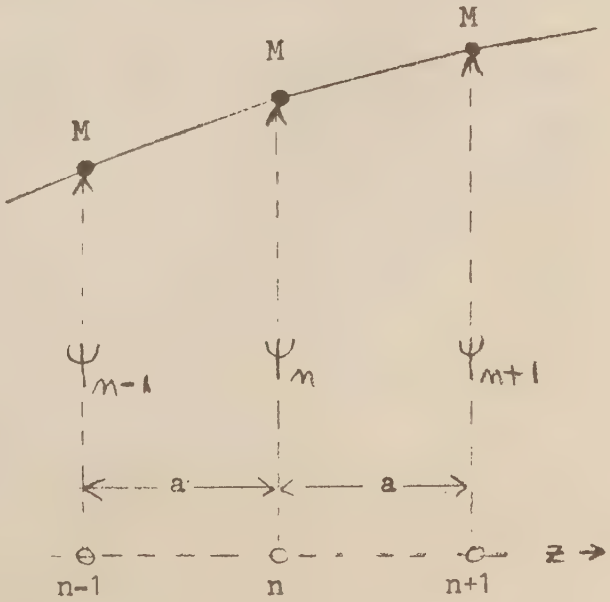


Fig. 1-12. Three successive beads on a uniform beaded string.

$$F_n = -(T/a)[(\psi_n - \psi_{n-1}) - (\psi_{n+1} - \psi_n)]$$

$$= -(T/a)[2\psi_n - \psi_{n-1} - \psi_{n+1}] \quad . \tag{84}$$

According to Newton's second law we therefore have

$$\frac{Md^2\psi_n(t)}{dt^2} = -(T/a)[2\psi_n - \psi_{n-1} - \psi_{n+1}] \quad . \tag{85}$$

Equation (85) is perfectly general, in the sense that it does not necessarily correspond to motion in a given mode, but can correspond to a general motion, i.e., to an arbitrary superposition of modes. But we now know that the easiest way to study the general motion is to study the individual modes. Thus we consider a single mode of angular frequency ω . In a mode we have

$$\psi_n(t) = A_n \cos \omega t \tag{86}$$

$$\frac{d^2\psi_n(t)}{dt^2} = -\omega^2 A_n \cos \omega t \tag{87}$$

where ω and the ratios $A_1 : A_2 : \cdots A_{n-1} : A_n : A_{n+1} : \cdots$ depend on the mode. Inserting Eqs. (86) and (87) into Eq. (85) and cancelling the common factor of $-\cos \omega t$, we obtain

$$M\omega^2 A_n = (T/a)[2A_n - A_{n-1} - A_{n+1}] \quad ,$$

i.e.,

$$A_{n-1} + A_{n+1} = [2 - (Ma/T)\omega^2]A_n \quad . \tag{88}$$

Equation (88) looks formidable. We have an unknown function A_n of an integer n satisfying the following condition: when you displace n first down by 1, and then up by 1, and then add the resulting functions, you get back the original function A_n, multiplied by a constant. How does one solve such a problem? By experience! You may (or someone did) notice a resemblance between Eq. (88) and the trigonometric identity

$$\sin(\theta - \theta_o) + \sin(\theta + \theta_o) = 2 \sin\theta \cos\theta_o \quad . \tag{89}$$

To utilize this analogy we need a dimensionless quantity θ (measured in radians) such that when n goes to $n+1$, θ goes to $\theta + \theta_o$. We know that z_n, the position along the z axis of the nth bead, increases by amount a when n increases to $n + 1$. However, z is not dimensionless but has dimensions of length. To get around this we simply "invent" a quantity k, whose physical significance we shall eventually understand (we hope), and which has dimensions rad/cm. Thus we define

$$\theta_n = kz_n \quad , \quad \theta_o = ka \tag{90}$$

(We could also add a constant, independent of n, but that is unnecessary.) We now try a solution of Eq. (88) of the form

$$A_n = A \sin\theta_n \quad , \tag{91}$$

where A is a constant, independent of n. Comparing Eq. (91) with Eq. (79), we see that if we find a solution of this form we can identify k

as the wave number. Comparing Eqs. (89) and (88), we see that Eq. (91) is a solution of Eq. (88) if and only if

$$2 \cos \theta_o = [2 - (Ma/T)\omega^2] \quad . \tag{92}$$

Rearranging Eq. (92) to use the identity $1 - \cos \theta_o = 2 \sin^2 \frac{1}{2} \theta_o$, setting $\theta_o = ka$, solving for ω^2, taking the square root, and dividing by k, we get the dispersion relation for the beaded string:

$$v(k) = \frac{\omega(k)}{k} = (Ta/M)^{\frac{1}{2}} \left[\frac{\sin (ka/2)}{(ka/2)}\right] \quad . \tag{93}$$

The motion of the bead is given by Eqs. (86) and (91):

$$\psi_n(t) = A \sin kz_n \cos \omega t \tag{94}$$

where k is related to ω by the second equality of Eq. (93).

Equations (93) and (94) are exact. They hold for a string with as few beads as one, or with as many as you wish. We have imposed the "boundary condition" that $\psi_n(t)$ is always zero at $z = 0$, but we could always remove that restriction by adding a constant to the argument kz_n of $\sin kz_n$. Similarly the boundary condition at the end $z = L$ is at our disposal. If we consider only violin strings fixed at $z = 0$ and L, then we shall want to retain $\sin kz_n$, and also we must have $\sin kL = 0$. (Remember the first and last beads are not at the ends of the string, but are separated from the ends by one string segment. Thus $z_n = na$, with

$z_o = 0$ at the fixed end.)

Let us check our solution, Eq. (94), by using it for the simplest possible case, where we have one bead. There is only one mode. The length of the string is $2a$, and this is one half wavelength, i.e., $2ak_1 = \pi$ (because $\frac{1}{2}$ wavelength is $\frac{1}{2}$ cycle, which is π radians), where k_1 means k for mode 1. The dispersion relation Eq. (93) is more conveniently written for present purposes)

$$\omega^2 = 4 \frac{T}{Ma} \sin^2 (ka/2) \quad . \tag{95}$$

Then $\sin^2(k_1 a/2) = \sin^2(\pi/4) = \frac{1}{2}$, and we find $\omega_1{}^2 = 2T/Ma$. This is what we would get by elementary considerations -- the factor of 2 is the result of having two strings. [See Eq.(P2), prob. 5.]

For more complicated cases (more beads), we always have that for N beads the length of the string is $(N+1)a$, which is always half of λ_1, the wavelength of the lowest mode. Thus we always have

$$k_1(N+1)a = \pi \, , \; k_2 = 2 k_1 \, , \; k_3 = 3 k_1 \, , \text{ etc.} \tag{96}$$

For each value of k, Eq. (93) gives ω.

Since knowing n means knowing $z = z_n$, we can always write Eq. (94) in a form that makes n disappear from the equation:

$$\psi_n(t) \equiv \psi(z,t) = A \sin kz \cos \omega t \quad . \tag{97}$$

Of course this function of z and t does not follow the string; it just

goes through the beads. But when the wavelength is large compared with the spacing a, i.e., for the low modes of a many-beaded string, $\psi(z,t)$ almost coincides with the string. (See Fig. 11.) In the limit as $a \to 0$, $\psi(z,t)$ coincides with the string, and then we can think of $\psi(z,t)$ as the displacement of the string at <u>any</u> point z. It is a continuous function, so that we can differentiate it and use all the lore of continuous functions. Our notation then becomes slightly modified. Instead of writing $d\psi_n(t)/dt$ we write $\partial\psi(z,t)/\partial t$, which is read "the partial derivative of ψ with respect to t, with z held constant" (we usually drop the phrase "with z held constant").

Let us examine the phase velocity $v(k)$ in the limit of long wavelength. We expand Eq. (93) in a Taylor's series about $ka/2 = 0$. We use [see Appendix, Eq. (4)]

$$\sin x = x - \frac{x^3}{3!} + \cdots ,$$

so that

$$\frac{\sin x}{x} = 1 - \frac{x^2}{3!} + \cdots \tag{98}$$

with $x = ka/2$. Then Eq. (93) becomes

$$v(k) = (Ta/M)^{\frac{1}{2}}\left[1 - \frac{(ka)^2}{24} + \cdots\right] \approx \left(\frac{T}{\rho}\right)^{\frac{1}{2}} , \text{ for } ka \ll 1 , \tag{99}$$

with $\rho \equiv M/a$, the average mass per unit length. We see that the velocity becomes independent of k, to the approximation that we can neglect $(ka)^2/24$ compared with unity. Thus for a continuous $(a \to 0)$, homogeneous (ρ = constant), flexible string, the wave velocity is independent of wave-

length. Such waves are said to be "non-dispersive," because in a superposition containing different wavelengths, all waves travel with the same velocity.

We don't want to strangle you with this string. Now that we know what a dispersion relation is, let us find one for a more interesting situation. Electromagnetic waves in the earth's ionosphere. We shall consider what happens when electromagnetic waves go through the earth's ionosphere. We shall find, as we already know by looking at the moon and stars, that visible light gets through very well. But we shall find that radio waves of frequencies less than 10 or 20 Mc do not get through. In fact, if they are emitted on earth, they bounce and return to earth. All of this follows from the dispersion relation -- plus whatever experimental data we may need on the density of free electrons per unit volume in the plasma which constitutes the ionosphere.

Later in this book we shall use Maxwell's equations to study electromagnetic radiation. Then we shall find, among other things, that the index of refraction $n(\omega)$ defined in Eq. (82) is given by

$$n^2(\omega) = \epsilon(\omega)\,\mu(\omega) \quad , \tag{100}$$

where $\epsilon(\omega)$ is the dielectric constant and $\mu(\omega)$ is the magnetic permeability of the medium. The dielectric constant is defined [see Volume 2, p. 323, Eq. (38)] by

$$\epsilon \underset{\sim}{E} = \underset{\sim}{E} + 4\pi \underset{\sim}{P} \tag{101}$$

where the polarization $\underset{\sim}{P}$ is the induced electric dipole moment per unit volume (see Volume 2, Sec. 9.8), and $\underset{\sim}{E}$ is the inducing electric field in the medium; $\underset{\sim}{P}$ is proportional to $\underset{\sim}{E}$, as we shall show below. Similarly, the magnetic permeability μ is defined in terms of the magnetization $\underset{\sim}{M}$, which is the induced magnetic dipole moment per unit volume, (see Volume 2, Sec. 10.8) and the inducing magnetic field $\underset{\sim}{B}$, as follows:

$$\frac{1}{\mu}\underset{\sim}{B} = \underset{\sim}{B} - 4\pi\underset{\sim}{M} \quad , \tag{102}$$

with $\underset{\sim}{M}$ proportional to $\underset{\sim}{B}$.

We have already discussed some properties of the ionosphere (see "plasma oscillations" in Sec. 1.2). It turns out that the induced magnetization $\underset{\sim}{M}$ is negligible; we shall take $\mu = 1$. Thus we need only consider the polarization $\underset{\sim}{P}$. The induced polarization is along (or opposite to) the total electric field $\underset{\sim}{E}$, where $\underset{\sim}{E}$ includes the induced field caused by the displacement of charge. Suppose $\underset{\sim}{E}$ is along x, and is given (at a certain fixed position) by

$$E_x = E_o \cos \omega t \quad , \tag{103}$$

where ω is the angular frequency of the electromagnetic radiation that is the source of the field $\underset{\sim}{E}$. The force on one free electron at that position in the plasma is partly due to E_x, and partly due to other causes (collisions with air molecules). These other forces average to zero, and we need not consider them.*

---

* You may think we should also include the force $q(\underset{\sim}{v}/c)X\underset{\sim}{B}$, where $\underset{\sim}{B}$ is the magnetic field due to the radiation. But it turns out that $|v/c|$ is so tiny we can neglect it.

We can write Newton's second law for one free electron having charge $q = -e$:

$$\frac{md^2x}{dt^2} = qE_x = qE_o \cos \omega t \quad . \tag{104}$$

The electron is driven at angular frequency ω , and of course it will oscillate at that frequency. (See driven oscillators, Volume 1, p. 211). Therefore we have $d^2x/dt^2 = -\omega^2 x$; inserting this in Eq. (104) gives

$$x(t) = -\frac{qE_x(t)}{m\omega^2} \quad . \tag{105}$$

The corresponding dipole moment per unit volume is equal to the number of free electrons per unit volume, N , times the dipole moment arising from the displacement of each, $qx(t)$. (See Volume 2, Sec. 9.8.) Thus Eq. (105) gives

$$P_x = Nqx(t) = -\frac{Nq^2E_x(t)}{m\omega^2} \quad . \tag{106}$$

According to Eqs. (101) and (106) the dielectric constant is given by

$$\epsilon(\omega) = 1 - \frac{4\pi Ne^2/m}{\omega^2} \quad . \tag{107}$$

From our earlier discussion of plasma oscillations [Eq. (28)] we have ω_p , the angular frequency for free oscillations of the electrons (called the plasma frequency or angular plasma frequency) given by

$$\frac{4\pi Ne^2}{m} = \omega_p^2 \quad . \tag{108}$$

Finally, combining Eqs. (100),(107), and (108), with $\mu = 1$, gives

$$n^2(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \quad . \tag{109}$$

An alternative way to write the dispersion relation Eq. (109) is to combine it with the relation $v = \omega/k = c/n$, which gives $k^2 = \omega^2 n^2/c^2$, or

$$k^2 = \frac{\omega^2 - \omega_p^2}{c^2} \quad . \tag{110}$$

Now that we have the answer, Eq. (109) or its equivalent Eq. (110), let us see if we can understand it. To do so we try various limiting cases, and see whether we can comprehend the physical behavior described by the formula.

Case 1: $\omega^2 \gg \omega_p^2$. If we can neglect ω_p^2 we see from Eq. (109) that we have $n \approx 1$, or equivalently, from Eq. (110), $\omega/k \approx c$. In this limit, the radiation pays no attention to the plasma. It goes right through with the same phase velocity ω/k as in vacuum, namely c. We know that this is true, because we can see the stars. Visible light has $\nu \sim 10^{15}$ cps, whereas $\nu_p \sim 10^7$ cps, according to our earlier discussion [see Eq. (26)]. The reason radiation of high frequency (compared with ν_p) goes through the ionosphere without being affected is as follows. The electric field of the radiation pushes on the free electrons. Their natural "response time" is given by the period of the free plasma oscillations. Thus the inertia of the electrons prevents them from responding immediately. If the driving field were a steady field, the electrons would eventually move until the field induced by their displacement cancelled the driving field to give

total field E_x of zero. We know that this is what happens for electrostatic fields in a conductor (the plasma is a conductor). But this cannot happen because the driving field has reversed before the electrons have time to move very much, provided ν is large compared with ν_p. Thus the electrons never get a chance to move so as to set up a cancelling field. The incident radiation is thus unaffected in the limit $\nu/\nu_p \to \infty$.

Case 2: $\omega^2 > \omega_p^2$, but ω_p^2 not negligible. We see from Eq. (109) that n^2 is less than 1, hence n is less than 1. That means that $v(k) = c/n$ is greater than c. Does this contradict the special theory of relativity? It had better not. If it does, something is wrong. Fortunately it doesn't. The velocity $v = \omega/k$ is a phase velocity. That is not the velocity at which signals travel. The theory of relativity only says that we cannot have a "causal" sequence of events propagating from one place to another at a velocity greater than c. But in a wave which has a time dependence $\cos \omega t$ at one place, and the same time dependence, $\cos \omega t$, (with perhaps a different amplitude) at another place, it is impossible to tell which oscillation at one place is the "cause" of the oscillation at the other place. This is because $\cos \omega t$ is periodic, and one cycle is just like any other. Such a time dependence is the result of a steady state that has been reached after all the "transients" in the driven oscillators have disappeared. If the "transmitter" suddenly increases the amplitude of its radiated wave, the complete system will take some time to reach a new steady state. Such a sudden increase in field at a transmitter will later cause a sudden change of the receiver's amplitude at its location far from the transmitter. This transient effect can be used to send a signal. If such a transient were to propagate faster than the velocity of light in vacuum,

c , it would indeed contradict relativity theory. However, it doesn't. Transients propagate not at the phase velocity ω/k , but at the group velocity, v_g which is given (as will be shown in Chap. 2) by

$$v_g = \frac{d\omega}{dk} \quad . \tag{111}$$

Using Eq. (110) we have

$$v_g^{-1} = \frac{dk}{d\omega} = \frac{\omega}{kc^2} = \frac{v}{c^2} \quad ,$$

i.e.,

$$v_g v = c^2 \quad . \tag{112}$$

We see that since the phase velocity $v = \omega/k$ is greater than c , the group velocity v_g is less than c . Relativity is saved!

Case 3: $\omega^2 < \omega_p^2$. Exponential attenuation in a reactive medium. If ω^2 is less than ω_p^2 we see that the square of the index of refraction as given by Eq. (109) is negative. What does that mean? Alternatively, from Eq. (110), k^2 is negative. This means that k is a pure imaginary quantity. What does that mean?

In order to understand what negative k^2 means, it helps to go back to the case where it is positive, when we thought we understood it (as being the square of k , which is 2π over the wavelength λ , which we do understand). We are dealing with continuous media (the graininess due to the individuality of the electrons is unimportant). Therefore we are nominally dealing with standing waves in the plasma of the form of Eq. (97),

$$\psi(z,t) = A \sin kz \cos \omega t \quad . \tag{113}$$

Actually we are at the moment more interested in traveling waves, which we have not yet studied officially. (We study them officially in Chap. 2.) Traveling waves of a definite frequency have the form (for travel in the $+z$ direction)

$$\psi(z,t) = A \cos (kz - \omega t) \tag{114}$$

(or, just as well, $\sin(kz - \omega t)$; as usual, we suppress uninteresting phase constants). This form is different from that of Eq. (113). But at the moment we are only interested in k and z , i.e., in the space dependence. If we hold t fixed and look at the space dependence at a given instant -- say $t = 0$ -- we have in Eqs. (113) and (114) the same space dependence, namely, an oscillation in space of the form

$$\varphi(z) = A \cos (kz + \alpha) \tag{115}$$

(where we include the constant α so as to include either $\sin kz$ or $\cos kz$) .

The function $\varphi(z)$ satisfies the differential equation

$$\frac{d^2\varphi(z)}{dz^2} = - k^2\varphi(z) \quad . \tag{116}$$

But let us now regard the differential equation Eq. (116) as more "fundamental" than its solution, Eq. (115). (It is really just a case of the chicken and the egg; it depends on your point of view.) We suppose that we have somehow

arrived at Eq. (116) with a theory that tells us how to find k^2, once ω^2 is known. If the theory tells us that k^2 is positive, we know then that $\varphi(z)$ has the form of Eq. (115). If it tells us that k^2 is negative, that is nothing to worry about. In that case we simply write, to avoid confusion,

$$k^2 = -|k|^2 \tag{117}$$

and rewrite Eq. (116) as

$$\frac{d^2\varphi(z)}{dz^2} = |k|^2 \varphi(z) \quad . \tag{118}$$

Equation (118) has the well-known solutions

$$\varphi(z) = e^{-|k|z} \quad \text{and} \quad \varphi(z) = e^{+|k|z} \quad . \tag{119}$$

Depending on the physical situation (i.e., where the transmitters are, which direction is the $+z$ direction, etc.), it may be necessary to use either one or the other or both of Eqs. (119).

In the present example of radio waves in the ionosphere with frequency below the "cutoff" frequency ω_p, our physical intuition tells us which solution we must use. The electrons move to new positions so as to produce a field which cancels the incident field. However, it isn't all done by the first electron the incident field encounters, so to speak. Instead the incident wave is exponentially attenuated with distance into the plasma. Thus, instead of Eq. (114) we find

$$\psi(z,t) = Ae^{-|k|z} \cos \omega t \quad , \tag{120}$$

where, in the present example, $\psi(z,t) = E_x(z,t)$ which we had previously considered only at a fixed z. (We shall show in Chap. 3 that in an electromagnetic wave the fields $\underset{\sim}{E}$ and $\underset{\sim}{B}$ are transverse to the propagation direction. That is why we take $\underset{\sim}{E}$ along $\hat{x}$ for propagation along $\hat{z}$.)

With this solution, the incident traveling wave eventually disappears. Where does the energy carried by the incident traveling wave go? Does its energy go into heat motion of the electrons? The answer is that it is reflected back towards the source. The behavior of the ionosphere towards radio waves of less than 10 Mc or so (the cutoff frequency -- it varies from day to day, and especially from day to night) is quite similar to the behavior of a silvered piece of glass -- a mirror -- to visible radiation. In each case the free electrons are set to vibrating by the incident field. The electrons themselves radiate in such a way as to cancel the forward propagating wave by destructive interference. In so doing they radiate a backward propagating wave -- the reflected wave.

To check this hypothesis about how a mirror works, you can estimate the number of free electrons per unit volume in silver (for example) by assuming there is one valence electron that becomes "free" per silver atom. Use the formula for ω_p and the fact that visible light has wavelength $\lambda \approx 5 \times 10^{-5}$ cm (from which you can calculate its frequency) to find whether the frequency for visible light is indeed less than the cutoff frequency ν_p for silver. (See problem 14.)

Surface waves on water. At equilibrium the surface of a body of water is flat and horizontal. When a wave is present there are two kinds of restoring force that tend to flatten the wave crests and restore equilibrium: one is gravity, the other is surface tension. For long wavelengths (more

than a few centimeters) gravity dominates. For millimeter wavelengths surface tension dominates. A very nice demonstration of circular surface-tension standing waves can be made as follows: Fill a paper or styrofoam cup up to the brim and then add a little more, so that the water bulges above the top of the cup. Tap the cup gently. Watch! A lovely experiment is to set the cup on a table or something and hold a spinning toy gyroscope against the table. Vibrations of the gyroscope drive the circular modes and, as the gyro slows down, you can see the wave length change, see resonances excited and die away, etc. To see the surface easily look at reflected sky. Alternatively, use a small bright light held several feet above the surface, and look at the pattern on the bottom due to the lens-like effect of the waves. The lowest gravitational mode in a closed body of liquid is sometimes called the "sloshing" mode. It is easily excited, as anyone knows who has tried to carry a pan of water without sloshing. Lake Geneva sometimes is excited into the sloshing mode (which is called a "seiche"). The period for this mode (in Lake Geneva) is about an hour. (See problem 16.)

Traveling waves in water are perhaps more familiar. Ocean waves coming in to shore and circular waves expanding outwards from a disturbance on a pond are examples. You can make nice traveling waves in a pan of water by blowing through a soda straw to excite the surface. Use a single light for illumination and blow near one end of the pan. Even if you blow steadily, the disturbance will be periodic. Alternatively, you can control the period by "tongueing" the straw.

Because of the great incompressibility of water, the excess of water that appears in a wave crest must flow in from the neighboring trough regions. Individual water drops in a water wave therefore undergo a motion that is a combination of longitudinal motion (forwards and backwards along

the wave propagation direction) and transverse (up and down) motion.

If the wavelength is small compared to the equilibrium depth of water we have what are called deep-water waves. The individual water droplets move in circles. A floating duck (or a droplet at the surface) undergoes a uniform circular motion with radius equal to the amplitude of the harmonic wave and with a period equal to that of the wave. On the crest of a traveling wave the duck has its maximum forward velocity; in a trough it has its greatest backward velocity. Water droplets below the surface travel in smaller circles, the radius of gyration decreasing exponentially with depth. The motion is negligibly small a few wavelengths below the surface.

The dispersion relation is given approximately by

$$\omega^2 = gk + \frac{T}{\rho} k^3 \quad , \qquad (121)$$

where $\rho \approx 1.0$ gm/cm^3 and $T \approx 72$ dyne/cm (surface tension) for water; $g = 980$ cm/sec^2. We do not advise trying to derive Eq. (121); it is difficult.

Esoteric examples. De Broglie's hypothesis, that a particle of momentum p has a wave number k given by $p = \hbar k$, along with the "Bohr frequency condition," that a particle of energy E has a wave frequency ω given by $E = \hbar \omega$, leads to a dispersion relation between ω and k, given the relation between E and p. Examples are given in Advanced Topic 2.

Problems, Chapter 1.

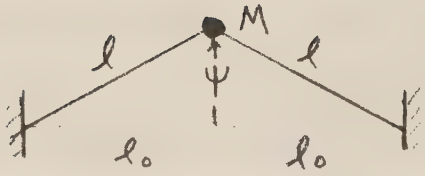
1. Use Eqs. (9) and (10) to decide whether for finite oscillation amplitude a pendulum's period is slightly longer or shorter than is implied by Eq. (11). Let the pendulum be a simple pendulum, i.e., one which satisfies Eq. (9) exactly.

2. If you set a small block of wood (or something) on a record-player turntable and look at it from the side as the turntable goes around, using only one eye so as to get rid of your depth perception, the apparent motion (i.e., motion projected perpendicular to your line of sight) is "harmonic," i.e., of the form $x = x_0 \cos \omega t$. (a) Prove that statement. (b) Make a simple pendulum by hanging a small weight (like a nut or bolt) from a string and hanging it from the back of a chair. Adjust the length of string until you can get your pendulum to swing in synchronization with the projected motion of the block on the turntable, with the record player set at 78 rpm. In this way you have a nice demonstration of the fact that the projection of a uniform circular motion is a harmonic oscillation. Also it is a nice way to measure g . If g has the standard "textbook value" of 980 cm/sec^2 , what should ℓ be? (for 78 rpm).

3. We have not discussed damping of harmonic motion (but see Volume 1, p. 207). Here is an interesting demonstration. Procure a raw egg, and a hardboiled egg. Use the eggs as bobs for two pendulums (a paper cup may work as a sling). Make the string lengths the same for each pendulum -- reasonably short, 20 cm or so. Set both pendulums going at the same time, with the same amplitude. Then wait and watch. What happens? Why? How?

4. This "home experiment" requires a hacksaw blade ($0.25 or so at any hardware store), a C-clamp (slightly more), a TV set, and a piano. Also a chair. The light emitted by a TV screen makes a nice stroboscope. The screen is actually dark "most of the time," and is on a small fraction of the time at a regular repetition rate ν_{TV}. The object is to measure ν_{TV}. Clamp the hacksaw blade to the chair with 2 to 4 inches extending. Illuminate with the TV (at night, with other lights out). Twang the blade and adjust the length until you can make the blade appear nearly to "stand still." Listen to the pitch. Find the pitch on the piano. Look up the pitch of that note (for a piano that is in tune) in the Handbook of Chemistry and Physics (indexed under "musical scales"). That's the answer. We'll tell you that it is either 60 or 120 cps. If the piano is not out of tune by an entire octave (a factor of 2 in frequency) you should be able to tell which it is.

5. Consider a stretched string with a single bead on it. The transverse displacement of the bead is $\psi(t)$. For $\psi = 0$ the two string segments are of length ℓ_o and the string tension is T_o. For $\psi \neq 0$ the two strings are each stretched from length ℓ_o to ℓ. Suppose that the string has the property that the tension T is given exactly by $T = T_o + C(\ell - \ell_o)$. (More generally, we could write T in a Taylor's series expansion at $\ell = \ell_o$, $T = T_o + C_1(\ell - \ell_o) + C_2(\ell - \ell_o)^2 + \ldots$. Instead we assume that only T_o and C_1 are different from zero.)



(a) Show that the exact non-linear differential equation governing the motion of M is

$$\frac{Md^2\psi(t)}{dt^2} = -2C\psi - 2(T_o - C\ell_o)\frac{\psi}{(\ell_o^2 + \psi^2)^{\frac{1}{2}}} \qquad (P1)$$

(b) Show that if the strings are perfect "Hooke's law" springs obeying a linear return force with spring constant C, and if their unstretched length (not ℓ_o!) is zero, then Eq. (P1) reduces to a linear equation, even for arbitrarily large displacements ψ.

(c) "Linearize" Eq. (P1) by expanding $(\ell_o^2 + \psi^2)^{-\frac{1}{2}}$. [Use Appendix Eq. (20), with $n = -\frac{1}{2}$ and with $x = \psi^2/\ell_o^2$], collecting like powers of ψ, and finally neglecting ψ^2 and higher powers of ψ.

Answer:

$$\frac{Md^2\psi}{dt^2} = -\frac{2T_o\psi}{\ell_o} . \qquad (P2)$$

(d) See if you can choose the parameters T_o, C, and ℓ_o so that the non-linear equation (P1) is identical with Eq. (30) for the simple pendulum, up through terms of order ψ^3. Use the expansion you found in part (c).

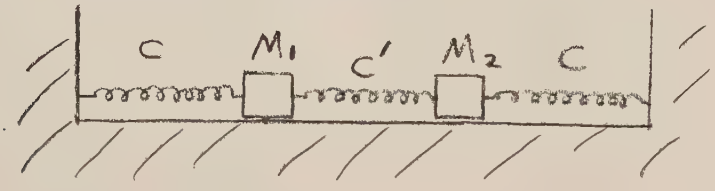
6. Express the initial transverse displacements and velocities of beads a and b at time $t = 0$ in terms of the four constants A_1, ϕ_1, A_2, and ϕ_2 of Eqs. (44). Conversely, solve for these four constants in terms

of the initial displacements and velocities. Take A_1 and A_2 to be positive. A phase constant, ϕ_1 for example, is said to be completely known if we know both its sine and cosine. Knowing its tangent is not sufficient (why?).

7. Find the oscillation frequencies for the normal modes of the coupled oscillators of Fig. 9. Answer: $\omega_1^2 = C^{-1}/(L + |M|)$, $\omega_2^2 = C^{-1}/(L - |M|)$. One mode has $I_a = I_b$, the other has $I_a = -I_b$. The correspondence with ω_1^2 or ω_2^2 depends on whether M is positive or negative.

8. Find the modes and their frequencies for the coupled springs and masses sliding on a frictionless surface shown below. At equilibrium the springs are relaxed.

(a) Take $M_1 = M_2 = M$. (b) Take $M_2 = 2M_1 = 2M$.



9. For the coupled pendulums of Fig. 8, let the masses be M_1 and M_2. Find the modes and their frequencies. [For $M_1 = M_2$ your answers should reduce to Eqs. (47) and (48).]

10. Consider the energy transfer between two weakly-coupled identical oscillators (Sec. 1.5). At $t = 0$, when a has all the oscillation energy and b has none, it is easy to see which is the "driven" oscillator (it is b) and which the "driving force" (it is a). Now consider time $t = \frac{1}{4} T_{beat}$,

a quarter of a beat cycle after $t = 0$. At that time pendulum a has lost half its energy, and b has gained that energy. Both pendulums have the same oscillation amplitude. How do they "know" which one is the driven and which the driver? How do they know which way the energy should keep flowing? Putting it differently, suppose you are allowed to look at the system and follow it through one oscillation (a fast oscillation of frequency about ω_1 or ω_2) at a time when both have the same energy. How can you predict whether the energy partition will (a) stay the same (b) change so as to flow to b (c) flow the other way. Try not to use the formulas. That is too easy. Look at the system itself -- what pulls on what, when, etc. Hint: the phase relationships are crucial.

11. Devise a damping mechanism ("friction") that will damp only mode 1 of the coupled pendulums of Fig. 8. Devise another that will damp only mode 2. Notice that friction at the supports(hinges) damps <u>both</u> modes. So does air resistance. These will not work.

12. Use Eqs. (95) and (96) to find the angular frequencies for the modes of a beaded string with two beads, with the string fixed at each end. For the answers see Eqs. (42) and (43).

13. Plot ω as a function of ka (ka going from zero to π) from the dispersion relation Eq. (95). Does this relation hold for ka greater than π ? (Give a physical interpretation of the answer.)

14. Calculate (roughly) the cutoff frequency ν_p for solid silver, assuming one free electron per atom (valence electron). Compare this with the

frequency of visible light, to see whether you would expect visible light to reflect from a silvered surface (if you had never looked in a mirror). What is the exponential decay length $|k|^{-1}$ for penetration into the silver? (See the discussion of electromagnetic waves in the ionosphere.)

15. Find the normal modes for three coupled pendulums. Choose equal masses, separations, lengths, and spring constants.

16. Sloshing mode in a pan of water, and seiches in Lake Geneva (and other lakes).

Fill a rectangular pan partly full of water. Push the pan a little. It sloshes. The best method is to lay the pan on a flat horizontal surface, fill it to the brim, and then overfill it so that it bulges above the level of the brim. Gently nudge the pan. After higher modes have damped out you will be left with the sloshing mode which will oscillate with very little damping. (It is a gravitational mode, even though you are using surface tension to hold the water "above the walls" -- so as to minimize damping.) The water level remains practically flat. (It is flat after the higher modes have damped out.) Assume it is flat throughout the motion, being horizontal when it passes through the equilibrium position, and tilted at the extremes of its oscillation. Let x be along the wave propagation direction and y be vertical upwards. Let $\bar{x}$ and $\bar{y}$ be the horizontal and vertical positions of the center of gravity of the water. Let $\bar{x}_o$ and $\bar{y}_o$ be the equilibrium values of $\bar{x}$ and $\bar{y}$. Find a formula for $\bar{y} - \bar{y}_o$ as a function of $\bar{x} - \bar{x}_o$. (A convenient parameter to work with is the water level at one end of the pan, relative to its equilibrium level.) The potential energy increase of the body of water is $mg(\bar{y} - \bar{y}_o)$. You will find that $\bar{y} - \bar{y}_o$ is proportional to $(\bar{x} - \bar{x}_o)^2$. Thus the center

of gravity has a potential energy like that of a harmonic oscillator. Use Newton's second law on the entire mass m, as if it were at the center of gravity. Find a formula for the frequency.

Answer: $$\omega^2 = 3 g h_o / R_o^2 , \qquad (P3)$$

where h_o is the equilibrium depth of the water, $g = 980$ cm/sec^2, and R_o is the half length of the pan along the direction of wave propagation -- along x. Try out this formula on your experiment with the pan of water, i.e., measure ω, h_o, and R_o and see how they agree with Eq. (P3).

According to an encyclopedia the average depth of Lake Geneva is about 150 meters. The length is about 60 kilometers (including the narrower western end). If we approximate the lake by a rectangular pan we can use Eq. (P3). What does Eq. (P3) predict, under those assumptions, for the period of the seiches that go in the long direction of the lake? (The observed period is of the order of an hour.) The seiches are probably excited by sudden differences in atmospheric pressure at one part of the lake relative to that at another. The observed amplitudes are up to five feet. In June 1954, a seiche of amplitude about ten feet occurred in Lake Michigan. It swept away a number of people fishing from piers.

17. What is the period of a simple pendulum of string length 1 meter?

18. Suppose one pendulum consists of a 1 meter string with a bob that is an aluminum sphere two inches in diameter. A second pendulum consists of a 1 meter string with a bob that is a brass sphere two inches in diameter. The two pendulums are set into oscillation at the same time and with the same amplitude A. After five minutes of undisturbed oscillation the aluminum pendulum is oscillating with one-half of its initial amplitude. What is

the oscillation amplitude of the brass pendulum?

Answer: The aluminum pendulum has amplitude 0.5 A (given). The brass pendulum has amplitude 0.88 A. Hints: The density of aluminum is 2.7 gm/cm^3; that of brass is 8.94 gm/cm^3. Why did the aluminum pendulum slow down? Will the brass one slow down as much?

19. A massless spring hangs from the ceiling. It has no mass attached to it. Its length is 20 cm. A mass M is now hung on the lower end of the spring. The mass is supported by your hand so that the spring remains relaxed. Suddenly you remove your supporting hand. The mass and spring oscillate. The lowest position of the mass during the oscillations is 10 cm below the place it was resting (when you supported it). What is the frequency of oscillation? What is the velocity when the mass is 5 cm below its original resting place?

Answers: 2.2 cps, 70 cm/sec.

A second mass of 300 gm is added to the first mass, to make a total of M + 300 gm. When this system oscillates it has half the frequency of the system with mass M alone. What is M ? Where is the new equilibrium position?

20. Suppose a given continuous physical system satisfies the dispersion relation

$$\omega^2 = A + Bk + Ck^2 \quad , \qquad (P4)$$

where A, B, and C are positive constants. Suppose that at infinite wavelength the group velocity goes to c (the velocity of light). What can you say about the constants A, B, and C? Now suppose that someone proposed a

dispersion relation

$$\omega = A + Bk + Ck^2 \quad . \tag{P5}$$

Show that such a dispersion relation violates the special theory of relativity.

21. In the derivation of the motion (transverse oscillations) of the beaded string [(Sec.1.6 near Fig.12 and Eq.(85)], suppose you have just reached Eq. (85). Suppose you have never heard of normal modes, so that you do not proceed to Eq.(86), which leads on to the exact solution and exact dispersion relation. You have heard of Taylor's series, so that you make the assumption that $\psi_n(t)$ can be approximated by a continuous function $\psi(z,t)$, where the continuous variable z at bead number n is z = constant + na (a is the bead spacing). Thus, $\psi_{n-1}(t) = \psi(z-a,t)$, $\psi_n(t) = \psi(z,t)$, and $\psi_{n+1}(t) = \psi(z+a,t)$. Expand ψ_{n+1} and ψ_{n-1} in a Taylor's series in the variable z , with t held constant. (The derivatives are then written as partial derivatives.) Carry terms through the fourth derivative of $\psi(z,t)$ with respect to z , with t held constant. Then express the right-hand side of Eq. (85) in terms of these derivatives. Write the left-hand side of Eq. (85) as the second partial derivative of $\psi(z,t)$ with respect to t , with z held constant. Show that Eq. (85) then becomes

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = \frac{2T}{Ma}\left\{-2\psi + 2\left[\psi + \frac{a^2}{2!}\frac{\partial^2\psi}{\partial z^2} + \frac{a^4}{4!}\frac{\partial^4\psi}{\partial z^4} + \frac{a^6}{6!}\frac{\partial^6\psi}{\partial z^6} + \cdots\right]\right\} \quad . \tag{P6}$$

If we take the first non-vanishing terms on the right hand side we obtain what is called the <u>classical wave equation</u>:

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = \frac{Ta}{M}\frac{\partial^2\psi(z,t)}{\partial z^2} \quad . \tag{P7}$$

Alternatively, we can see by inspection of Eq. (P6) that if $\partial^2\psi/\partial z^2$ is proportional to ψ, then the whole right-hand side of Eq. (P6) is proportional to ψ. (Show this.) Let us call this proportionality constant $-k^2$ if it is negative, and $+|k|^2$ if it is positive. Thus we assume either

$$\frac{\partial^2\psi}{\partial z^2} = -k^2\psi \quad , \tag{P8}$$

or, instead,

$$\frac{\partial^2\psi}{\partial z^2} = +|k|^2\psi \quad . \tag{P9}$$

Then the left side of Eq. (P6) is proportional to ψ. We assume the proportionality constant is $-\omega^2$. We then have, corresponding to Eq. (P8) (as you should show),

$$-\omega^2 = \frac{2T}{Ma}\left[-2 + 2\left\{1 - \frac{k^2a^2}{2!} + \frac{k^4a^4}{4!} - \frac{k^6a^6}{6!} + \ldots\right\}\right] \quad . \tag{P10}$$

But we have [Appendix Eq. (5)]

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{x!} - \ldots \quad .$$

Thus Eq. (P10) gives

$$\omega^2 = \frac{4T}{Ma}\left[1 - \cos ka\right] \quad . \tag{P11}$$

Similarly if we have Eq. (P9), you should show that we obtain

$$\omega^2 = \frac{4T}{Ma}\left[1 - \cosh |k|a\right] \quad . \tag{P12}$$

It is easy to show that Eq. (P11) is equivalent to Eq. (95), the dispersion relation we obtained exactly. The solutions for $\psi(z,t)$ are then any linear combination of the four functions $\sin\omega t \sin kz$,

$\sin \omega t \cos kz$, $\cos \omega t \sin kz$, and $\cos \omega t \cos kz$.

For Eq. (P9) it is easy to show that the solutions $\psi(z,t)$ are any linear combinations of the four functions $\sin \omega t\, e^{+|k|z}$, $\sin \omega t\, e^{-|k|z}$, $\cos \omega t\, e^{+|k|z}$, and $\cos \omega t\, e^{-|k|z}$. Solutions such as these correspond to reactive media. You have encountered this once [see the discussion near Eq. (118).]

The most fascinating thing about Eqs. (P9) and (P12) is that they are completely wrong! It turns out that for an ω^2 so great that Eq. (P11) cannot be satisfied (because $\cos ka$ would have to be less than -1), the beaded string takes on a zig-zag or "sawtooth" appearance and cannot be represented by a differentiable function.

Chapter 2 - Waves

Section		Page
2.1	Introduction.	1
2.2	Harmonic traveling waves.	5
2.3	Superposition of two harmonic traveling waves. Group velocity.	15
2.4	Superposition of many harmonic waves. Pulses and wave packets.	23
2.5	Wave equations.	40
	Problems.	59

2.1 Introduction.

In Chapter 1 we considered the normal modes of closed systems, i.e., systems having a definite number of moving parts and hence a definite number of degrees of freedom. In considering systems in the continuous approximation, there are effectively an infinite number of degrees of freedom, so we redefine a closed system to be one with definite spatial boundaries which confine the oscillating material. (In the case of electromagnetic waves there is no material, but the fields can be confined by mirrors at the boundaries.) We found that the steady-state oscillations of such a closed system are described by its modes, which we call standing waves when we have a continuous system.

In any real physical situation one does not deal with completely isolated (i.e., closed) systems. The energy stored in an almost-closed system gradually leaks away, one way or another. For example, the vibration energy in a plucked guitar string is dissipated after several hundred vibrations. Some of the energy goes into heating of the string. Most of the energy is "radiated" in the form of traveling sound waves which spread out from the string (and sounding board) into the essentially "open" system which consists of the air in the room. If the walls of the room are covered with sound-absorbing material, so that no waves are ever reflected back to the transmitter, then the room appears to the guitar like a completely open system with no outer boundaries. One important characteristic of an open medium (i.e., an open continuous system) is that it will carry harmonic traveling waves of any wavelength, because there are no boundaries to demand that the displacement be zero at some place (as at $z = L$ for the violin string). Consequently any frequency is allowed.

In general, a physical system may exhibit either standing waves or traveling waves or both. The same physical laws govern both types of waves; what sort of waves are present in a given system depends on the boundary conditions of the system. (For example, fixing the ends of a vibrating string allows only standing waves to be present.) The relation between wavelength and frequency for a given medium, the dispersion law, is the same for traveling waves as for standing waves in the same medium. That can be seen to be true for the case of the beaded string considered in Sec. 1.6. Our derivation of the dispersion law did not depend on boundary conditions. The derivation goes through in almost the same way if one looks for traveling-wave solutions of the form $A \cos (kz_n - \omega t)$ as it did when we looked for standing-wave solutions of the form $A \sin kz_n \cos \omega t$.

When an oscillating source is coupled to an open medium it radiates traveling waves which proceed out from the source into the medium. If the medium is not perfectly open some of the waves may be reflected back to the source. We shall usually avoid this complication. However, it is an extremely important consideration in practice. In order to experience it, sing a slowly ascending "siren" sound into a mailing tube or a large jar. You will undoubtedly find one or two "resonances." At a resonance you can feel a striking change in the feel of the "impedance" or "resistance to air flow" of your singing. This "feedback" can be so strong as to make it difficult to sing. An especially strong effect can be felt if you whistle across the top of a pop bottle. (The bottle should be partly filled so as to bring the lowest resonant mode of the bottle within comfortable whistling range.)

In Sec. 2.2 we shall discuss the properties of harmonic traveling waves

produced by a harmonically-oscillating transmitter coupled to a one-dimensional, open, homogeneous medium. We shall show how to express mathematically the fact that the harmonic traveling wave propagates without changing its shape. This means that the displacement $\psi(z,t)$ has the same value at the place z_2 at time t_2 as it had at the place z_1 at the earlier time t_1, when t_1 and t_2 are connected by

$$t_1 = t_2 - \frac{(z_2 - z_1)}{v(\omega)} \quad .$$

Therefore if you know ψ at $z = 0$ as a function of time t, it is easy to find it at $z > 0$ for $t > 0$. In Sec. 2.2 we derive the important relation for the phase velocity of a traveling harmonic wave:

$$v(\omega) = v(k) = \frac{\omega}{k} \quad .$$

In Sec. 2.3 we consider what happens when the transmitter oscillates in a superposition of two different harmonic oscillations with different frequencies. We find that the amplitude modulations do not propagate at the phase velocity $v = \omega/k$. Instead they propagate at the group velocity $(\omega_1 - \omega_2)/(k_1 - k_2)$. For the interesting case where ω_1 and ω_2 are so close together that we can neglect all but the first term in the Taylor's series expansion of this ratio, it becomes just $d\omega/dk$:

$$\text{group velocity} = \text{velocity of modulations} = \frac{d\omega}{dk} \quad .$$

In Sec. 2.4 we consider a superposition of many frequency components

occupying a narrow frequency band of width $\Delta\omega$. We find that in the limit that there is a huge number of frequency components in the band $\Delta\omega$, we can form a superposition $\psi(t)$ which has the property that it is large near $t = 0$, rapidly falls to small values,and never becomes large again. Such a function is called a pulse. By examining our example we reach the conclusion that for any pulse one has, approximately, the very important relation

$$\Delta\nu \, \Delta t \approx 1 ,$$

where $\Delta\nu$ is the frequency bandwidth and Δt is the time duration of the pulse.

Then we examine the wave group or wave packet that is radiated by a source whose time dependence is that of a pulse. The wave packet travels with the group velocity. We find that the length of the packet Δz at the time it is emitted is related to the range of wave number Δk present in the wave packet by

$$\Delta k \, \Delta z \approx \Delta\omega \, \Delta t \approx 2\pi ,$$

where the first (approximate) equality is obtained from the dispersion relation and the group velocity, and the second is our result above. We also find that as the packet propagates in the medium it "spreads," i.e., Δz gets larger. This is because different "sub-bands" of the band of frequencies $\Delta\omega$ have slightly different group velocities. It is perhaps obvious that when something depends on a "band within a band," it depends on higher derivatives of the dispersion relation, and indeed the spread

of the packet turns out to depend on $d^2\omega/dk^2$.

In Sec. 2.5 we consider wave equations, which are partial differential equations relating the space and time derivatives of wave functions $\psi(z,t)$ [or more generally, $\psi(x,y,z,t)$] and describing all the standing or traveling waves possible to the system. We show how you can find the differential equation if you know the dispersion relation. We also show for the very important example of electromagnetism how Maxwell's equations lead to a wave equation. We do this for vacuum and find (not to our surprise) that Maxwell's equations predict the occurrence of non-dispersive electromagnetic waves having propagation velocity c . We also consider a linear, homogeneous, isotropic dispersive medium and derive the result that the index of refraction $n(\omega)$ is given by

$$\frac{c}{v(\omega)} \equiv n(\omega) = \sqrt{\mu(\omega)\epsilon(\omega)} \quad .$$

We then examine the implication of "imaginary index of refraction," which we first encountered in discussing waves in the ionosphere.

2.2. Harmonic traveling waves.

In Sec. 1.6 we considered a simple example of a closed system consisting of a continuous homogeneous medium with definite boundary conditions, namely the uniformly-beaded string with fixed ends. We found the standing waves of the system. Now we wish to consider open systems. The open system that corresponds to the beaded string of Sec. 1.6 can be taken to be a semi-infinite beaded string stretching from $z = 0$ to $z = \infty$. The string is attached at $z = 0$ to the output terminal of some device ("transmitter") that can shake the

string and thus radiate traveling waves along the string. In contrast to the case of the closed system with boundary conditions, an open continuous system can carry waves of any frequency. We shall show that the relation between ω and k is the same for the traveling waves of the open system as for the standing waves of the closed system (provided one has the same medium, i.e., same string tension, same bead mass and spacing). We shall also show that, if the transmitter at $z = 0$ shakes the string harmonically with a transverse displacement $D(t)$ of the form

$$D(t) = A \cos \omega t \ , \tag{1}$$

then the traveling wave (called a "harmonic traveling wave") has the form

$$\psi(z,t) = A \cos (\omega t - kz) = A \cos (kz - \omega t) \ . \tag{2}$$

Notice that in writing Eq. (2) we have already satisfied the boundary condition at $z = 0$, namely that the string is attached to the transmitter; this we have

$$\psi(0,t) = D(t) = A \cos \omega t \ . \tag{3}$$

There is no other boundary condition to satisfy. This traveling wave is shown in Fig. 1, where we use the nomenclature $T(=2\pi/\omega)$ and $\lambda(2\pi/k)$ instead of ω and k.

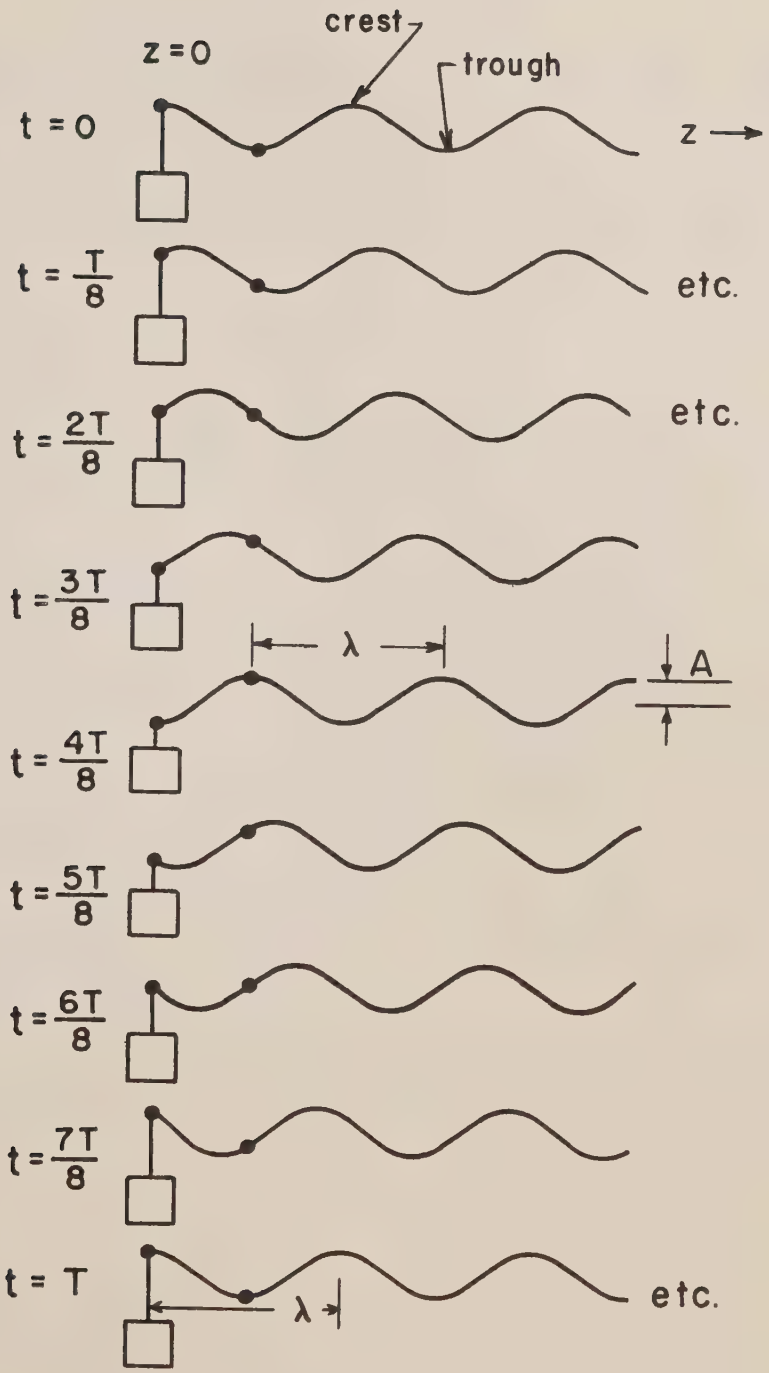


Fig. 2-1. Driving force at $z = 0$ describes harmonic motion of period T. Harmonic traveling wave propagates in $+z$ direction. The wavelength is λ. The velocity is λ/T. Every point on the string undergoes the same harmonic motion as that at $z = 0$, but at a later time.

Now we shall show that Eq. (2) is indeed the form of the traveling wave, and that the dispersion relation is the same as for the closed system. If you will reexamine our derivation in Sec. 1.6 of the form of the standing waves for the beaded string,

$$\psi(z,t) = A \sin kz \cos(\omega t + \phi) \quad , \tag{4}$$

You will find that we could just as well have found

$$\psi(z,t) = A \cos kz \cos \omega t \quad . \tag{5}$$

Equations (4) and (5) correspond to different boundary conditions at the ends of the string and different phase constants. The boundary conditions not only forced us to choose as we did between Eq. (4), Eq. (5), or some linear combination of them; they also demanded that only certain wavelengths be "allowed," and thus that only certain frequencies be allowed, namely those wavelengths and frequencies corresponding to the lowest mode, next highest, etc. Once the allowed wavelengths (i.e., values of k) were "selected" by the boundary conditions, the dispersion relation $\omega(k)$ give the frequencies. The dispersion relation <u>itself</u> did not impose any limitations on the frequencies.

We are now dealing with an open system. There are no boundary conditions [except the one given by Eq. (3)]. Therefore there is no restriction imposed on the choice between Eqs. (4) and (5) and no restriction on allowed values of k. Since there is no restriction, we are free to form superpositions to search for a wave form that satisfies our intuitive picture of a traveling wave (a picture perhaps formed by looking at traveling water waves). By trial and error we decide to set the phase constant ϕ in Eq. (4) at -90 deg, so that $\cos(\omega t + \phi)$ is just $\sin \omega t$. Then if we

superpose Eqs. (4) and (5), we obtain

$$\psi(z,t) = A\,[\cos kz \cos \omega t + \sin kz \sin \omega t] = A \cos (kz - \omega t) \ , \qquad (6)$$

which is the same as Eq. (2). [We could, of course, have just as well obtained $\sin (kz - \omega t)$. But that would not have satisfied the boundary condition Eq. (3).]

This "derivation" probably leaves you uneasy. We hope you will be more comfortable when you have satisfied yourself that Eq. (6) really does "look like" it should to accord with your intuitive picture. Also we shall call upon your intuition to recognize that for a given transmitter oscillation at $z = 0$, there must be a unique solution -- the same motion always produces the same kind of waves. Furthermore, the derivation which gave us Eqs. (4) and (5) and the dispersion relation could not have given any other solution except a linear combination of Eqs. (4) and (5) . Finally, no linear combination of Eqs. (4) and (5) except Eq. (6) can give a traveling wave satisfying the boundary condition at $z = 0$.

As usual, we are not really interested in strings. The importance of Eq. (6) is that it has the form taken by traveling waves in a vast number of interesting physical situations. All of these situations have common features:

1. The transmitter oscillates harmonically, as in Eq. (1). For example, instead of a string we may have an infinitely long cylindrical tube filled with air and driven at one end by a piston undergoing longitudinal simple harmonic motion. Waves of compression and rarefaction -- sound waves -- propagate down the tube. (The tube need not be infinitely long provided we terminate it with sound-absorbing material which will not reflect incident

sound waves.)

2. The medium is homogeneous. For example, if the mass per unit length of the string were not constant then the standing waves would not be of the form of Eqs. (4) or (5), but would have the form

$$\psi(z,t) = f(z) \cos \omega t \ , \qquad (7)$$

Where $f(z)$ would depend on the particular way in which the bead masses and spacings varied with z and on the boundary conditions (ends fixed or free). For this more complicated case one would not obtain traveling waves of the form of Eq. (6).

3. There is some boundary condition at $z = 0$ (i.e., at the transmitter), which depends on how the transmitter goes about radiating waves into the medium and how the transmitter is "coupled" to the medium. Even in the case of the string, we could have had a different condition at $z = 0$ than the one we assumed (i.e., that the string has the same displacement $D(t)$ as the "output hook" to which the string is attached). We could have attached the string to the transmitter through a spring, for example. (See problem 2.) Then there would have been a phase shift between the displacement $D(t)$ of the output hook and the displacement $\psi(0,t)$ of the wave at $z = 0$. When we study radiation of sound waves by a vibrating piston (or speaker cone), we shall have to examine the "physics" of what goes on at the transmitter in order to relate the amplitude and phase of the radiated wave to that of the transmitter. Similarly, when we study electromagnetic radiation, we shall examine the coupling between the moving electrons in the antenna system and the radiated waves of electric and magnetic fields.

4. The form of Eq. (6) holds only for one-dimensional waves. For example,

it holds for sound waves traveling down a tube. However, it does not hold for sound waves emitted by a small whistle suspended by a thread in the middle of a room with sound-absorbing walls. The room is indeed an effectively open medium. The whistle indeed vibrates with harmonic motion and emits harmonic traveling waves. But these waves propagate out in all directions, not just along z. They are called outgoing spherical waves. We shall not study them in any detail in this course, but we shall write down their form now, to satisfy your curiosity. (They are discussed further in Advanced Topic 5.) Instead of Eq. (6) one finds for an outgoing spherical harmonic wave

$$\psi(r,t) = A \frac{\cos(kr - \omega t)}{r} , \qquad (8)$$

where r is the distance from the whistle located at $x = y = z = 0$, i.e., r is $(x^2 + y^2 + z^2)^{\frac{1}{2}}$. (In writing Eq. (8) we have assumed that the source emits equally well in all directions, which we can arrange by using a rapidly expanding and contracting sphere as a source instead of a whistle).

The factor r^{-1} in Eq. (8) is interesting, since it implies that the magnitude of the displacement ψ decreases with increasing distance from the source. No such factor occurs in the one-dimensional traveling wave of Eq. (6). One can understand the factor r^{-1} in the spherical wave, and the constancy of amplitude in the one-dimensional wave in terms of energy conservation, as follows: We shall find (in Chap. 3) that the energy density of a traveling wave is proportional to the square of the displacement ψ^2. More precisely, for waves on a string, and for sound waves, the energy density in a traveling wave is proportional to the square of the velocity of the oscillation, $(\partial\psi/\partial t)^2$ (rather than to the square of the displacement,

ψ^2) . Thus for the one-dimensional traveling wave, Eq. (6), the energy contained between z and $z + dz$ at time t is given by

$$\text{energy in } dz = \text{const.} \left[\frac{\partial\psi(z,t)}{\partial t}\right]^2 dz \quad . \tag{9}$$

$$= \text{const.}\ \omega^2 A^2 \sin^2(\omega t - kz) \quad . \tag{10}$$

We see from Eq. (10) that if we follow the traveling wave (i.e., if we consider a point traveling so that the phase $\omega t - kz$ remains constant) the energy in interval dz (which also remains the same "thickness") is constant. In the case of the spherical waves, suppose that we assume that the displacement is given by $A \cos(\omega t - kr)$ times $f(r)$ an unknown function of r. Thus, assume

$$\psi(r,t) = A f(r) \cos(\omega t - kr) \quad . \tag{11}$$

The energy contained in a spherical shell of thickness dr and volume $4\pi r^2 dr$ is the energy density times the volume:

$$\text{energy in } dr = \text{const.} \left[\frac{\partial\psi(r,t)}{\partial t}\right]^2 4\pi r^2 dr \tag{12}$$

If we follow points in a spherical shell of constant phase $\omega t - kr$ as it expands with increasing time, the energy in the shell must remain constant. Therefore $f^2(r)r^2$ must be constant, i.e., $f(r)$ = constant times r^{-1}.

Now we shall restrict our attention to one-dimensional waves of the form of Eq. (6). Let us examine the properties implied by Eq. (6). The traveling wave $\cos(kz - \omega t)$ has the following features in common with the standing wave $\cos kz \cos \omega t$: (1) at fixed z the displacement ψ is a harmonic oscillation of the form $\cos(\omega t + \phi)$; (2) at fixed t the shape of the wave is a sinusoidal function of z of the form $\cos(kz + \alpha)$. The

moving parts in a traveling wave do not all pass through equilibrium simultaneously, in contrast to the standing wave, for which all moving parts (parts at all locations z) do pass through their equilibrium positions simultaneously. For the traveling wave there is a relative phase shift that corresponds to the time delay for a given value of displacement to travel from one place to another. The oscillations at locations $z = 0$ and z are related as follows: The displacement at time t at place z is the same as the displacement at $z = 0$ at the earlier time t', where

$$t' = t - \frac{z}{v} \quad , \tag{14}$$

and where v is the velocity of the traveling wave.

Thus, if the displacement at $z = 0$ at time t' is given by

$$\psi(0,t') = A \cos \omega t' \quad , \tag{15}$$

the displacement at z at the later time $t = t' + (z/v)$ is given by substituting Eq.(14) into Eq. (15) to obtain

$$\psi(z,t) = A \cos \omega(t - \frac{z}{v}) \quad . \tag{16}$$

By comparison of Eqs. (16) and the form $A \cos (\omega t - kz)$ [$= A \cos (kz - \omega t)$] of Eq. (6) we see that k is ω/v, i.e., the "phase velocity" $v = v(k)$ is given by the formula

$$\boxed{v(k) = \frac{\omega(k)}{k}} \quad . \tag{17}$$

This may also be written in the alternate forms

$$v = \lambda\nu \; ; \; v = \frac{\lambda}{T} \quad , \tag{18}$$

by using the relations $k = 2\pi/\lambda$, $\omega = 2\pi\nu$, and $\nu = T^{-1}$.

Equation (17) [and the equivalent form, Eq. (18)] is so important that we now give an alternative derivation. We define the phase function $\phi(z,t)$ of a harmonic traveling wave propagating in the $+z$ direction to be the argument of the wave function $\cos(\omega t - kz)$,

$$\phi(z,t) = \omega t - kz \quad . \tag{19}$$

[We have suppressed a possible phase constant in Eq. (19).] At a given z the phase increases linearly with time, as ωt. At a given time t, the phase decreases linearly with z, as $-kz$. Going to greater z decreases the phase, because it corresponds to waves emitted at earlier times. (Our sign convention for positive phase is not universal. Some people prefer to call $kz - \omega t$ the phase.) If we wish to follow a given wave crest [maximum of $\cos\phi(z,t)$] or trough [minimum of $\cos\phi(z,t)$], we must keep looking at different z for different t, so as to keep the phase $\phi(z,t)$ constant. Thus by taking the total differential of $\phi(z,t)$ and setting the result equal to zero, we can find the relation between z and t for a point of constant phase: The total differential of $\phi(z,t)$ is given by

$$d\phi = \left(\frac{\partial\phi}{\partial t}\right) dt + \left(\frac{\partial\phi}{\partial z}\right) dz = \omega\, dt - k\, dz \quad , \tag{20}$$

which is zero provided dt and dz are related by

$$v = \left(\frac{dz}{dt}\right)_{[d\phi=0]} = \frac{\omega}{k} \quad , \tag{21}$$

which is Eq. (17).

2.3 Superposition of two harmonic traveling waves. Group velocity.

Suppose that the displacement $D(t)$ of the transmitter at $z = 0$ does not have the simple harmonic time dependence $D(t) = A \cos \omega t$ [Eq. (1)], but has a more complicated dependence $D(t) = f(t)$. It turns out that a wide class of functions $f(t)$ can be expressed as linear superpositions (sums) of harmonic functions of the form $A(\omega)\cos [\omega t + \phi(\omega)]$, where the amplitude $A(\omega)$ and the phase constant $\phi(\omega)$ depend on ω and of course on the function $f(t)$ that is to be expressed as a superposition. You will study (or perhaps have studied) how to determine the amplitudes $A(\omega)$ and phase constants $\phi(\omega)$ by the methods of Fourier analysis. Here we shall consider only a superposition containing two terms. That will be sufficient to give us very interesting results that will lead us eventually to an understanding of how a "pulse" or "wave group" propagates in a dispersive medium (one where the phase velocity depends on wavelength).

Let us assume the transmitter at $z = 0$ oscillates in a superposition of two harmonic motions with angular frequencies ω_1 and ω_2. We shall not miss any interesting results if we let the amplitudes and phase constants of the two contributions be the same. Therefore we assume we have for the oscillating element (output hook) of the transmitter

$$D(t) = A \cos \omega_1 t + A \cos \omega_2 t \quad . \tag{22}$$

The string is attached to the transmitter in such a way that at $z = 0$, $\psi(z,t)$ is given by

$$\psi(0,t) = D(t) \quad . \tag{23}$$

Because of the linearity of the waves, i.e., because of the fact that they satisfy the superposition principle, the two contributions to the transmitter

motion $D(t)$ will radiate traveling waves independently. Thus the wave $\psi(z,t)$ will be just the superposition of the two harmonic traveling waves $\psi_1(z,t)$ and $\psi_2(z,t)$ that would be present if one or the other of the transmitter oscillations $A \cos \omega_1 t$ or $A \cos \omega_2 t$ were present by itself. From our previous discussion [Eqs. (1) and (2)], ψ_1 and ψ_2 are given by

$$\psi_1(z,t) = A \cos (\omega_1 t - k_1 z) \ , \tag{24}$$

$$\psi_2(z,t) = A \cos (\omega_2 t - k_2 z) \ . \tag{25}$$

The traveling wave $\psi(z,t)$ is not a harmonic traveling wave; it is the superposition

$$\psi(z,t) = \psi_1(z,t) + \psi_2(z,t) = A \cos(\omega_1 t - k_1 z) + A \cos(\omega_2 t - k_2 z) \tag{26}$$

of two harmonic traveling waves with <u>different frequencies</u>. (If ω_1 and ω_2 are equal then the superposition, Eq. (26), can be expressed in the form $A' \cos (\omega t - kz)$ and is thus a harmonic traveling wave. See problem 3.) Each of the harmonic-traveling-wave contributions ψ_1 and ψ_2 travels with its own phase velocity:

$$v_1 = \omega_1/k_1 \ , \quad v_2 = \omega_2/k_2 \ , \tag{27}$$

where in each case ω and k are related by the dispersion relation $k(\omega)$ for the medium. Thus the wave $\psi(z,t)$ has a very complicated appearance in general.

<u>Velocity of propagation of beats</u>. It is very instructive to consider the special case [of Eqs. (24) and (25)] where the beat frequency $\omega_1 - \omega_2$ is small compared with the average frequency ω_{av} :

$$\omega_b = |\omega_1 - \omega_2| \ ; \ \omega_{av} = \tfrac{1}{2}(\omega_1 + \omega_2) \ ; \ \omega_b \ll \omega_{av} \ . \tag{28}$$

Then the oscillation of the transmitter output exhibits beats (Sec. 1.5), as we see by writing Eq. (22) in the equivalent form [Sec. 1.5, Eq. 61a)]

$$D(t) = [2A \cos \tfrac{1}{2}(\omega_1-\omega_2)t] \cos \omega_{av}t \equiv A(t) \cos \omega_{av}t \ , \qquad (29)$$

which is plotted in Fig. 10, Chap. 1. Thus the wave $\psi(z,t)$ at $z = 0$ has the form

$$\psi(0,t) = D(t) = [2A \cos \tfrac{1}{2}(\omega_1-\omega_2)t] \cos \omega_{av}t \ . \qquad (30)$$

Now comes the interesting part. We want to find the corresponding "beat" form of Eq. (26). We use the trigonometric identity

$$\cos x + \cos y = 2 \cos \tfrac{1}{2} (x - y) \cos \tfrac{1}{2} (x + y) \ ,$$

with $x = \omega_1 t - k_1 z$ and $y = \omega_2 t - k_2 z$, to obtain from Eq. (26),

$$\psi(z,t) = A \cos (\omega_1 t - k_1 z) + A \cos (\omega_2 t - k_2 z) \ ,$$

the equivalent form

$$\psi(z,t) = \left\{2A \cos \tfrac{1}{2} [(\omega_1-\omega_2)t - (k_1-k_2)z]\right\} \cos (\omega_{av}t - k_{av}z) \ , \qquad (31)$$

with

$$k_{av} = \tfrac{1}{2} (k_1 + k_2) \ , \quad \omega_{av} = \tfrac{1}{2} (\omega_1 + \omega_2) \ . \qquad (32)$$

Equation (31) is exact, but its most interesting realization is for $\omega_1 - \omega_2 \ll \omega_{av}$. The factor

$$\psi \text{ ("carrier wave")} \equiv \cos (\omega_{av}t - k_{av}z) \qquad (33)$$

has the "fast" frequency ω_{av} and the "short" wavelength $2\pi/k_{av}$. (The term "carrier wave" is taken from the radio engineer's jargon.) The carrier wave propagates at the average phase velocity defined by

$$v_{av} \equiv \omega_{av}/k_{av} \quad . \tag{34}$$

The factor

$$A(z,t) \equiv 2A \cos \tfrac{1}{2} [(\omega_1 - \omega_2)t - (k_1 - k_2)z] \tag{35}$$

is an amplitude-modulation wave. At $z = 0$, $A(z,t)$ has the amplitude-modulation time dependence

$$A(0,t) = 2A \cos \tfrac{1}{2} (\omega_1 - \omega_2)t \quad , \tag{36}$$

which is just the time variation of amplitude that corresponds to the beats of Eq. (29). At any other fixed z, $A(z,t)$ as given by Eq. (35) again gives a time dependence corresponding to beats, but with a phase shift corresponding to a time delay for the beat to propagate from $z = 0$ to z. The beat does not propagate with the phase velocity $v_1 \approx v_2 \approx v_{av}$. The phase function $\phi_{beat}(z,t)$ for the beat cycle is given by

$$\phi_{beat} = \left[(\omega_1 - \omega_2)t - (k_1 - k_2)z \right] \quad . \tag{37}$$

The beats propagate with a velocity v_{beat} such that ϕ_{beat} remains constant; that is, to "follow" a beat we must maintain a relation between z and t such that ϕ_{beat} does not change. Thus, proceeding in a manner analogous to that used in Eqs. (20) and (21) we find the total differential

$$d\phi_{beat} = (\omega_1 - \omega_2)dt - (k_1 - k_2)dz \quad . \tag{38}$$

The condition that ϕ_{beat} be constant is $d\phi_{beat} = 0$, which gives

$$v_{beat} = \left(\frac{dz}{dt}\right)_{(d\phi_{beat}=0)} = \frac{(\omega_1 - \omega_2)}{(k_1 - k_2)} \quad . \tag{39}$$

Equation (39) is exact. We can put it in a more cogent form for the interesting case $\omega_1 - \omega_2 \ll \omega_{av}$. We assume that we know the dispersion relation $\omega = \omega(k)$. Then $\omega_1 = \omega(k_1)$, $\omega_2 = \omega(k_2)$. Let us expand $\omega(k)$ in a Taylor's series about $k = k_2$. Then at a neighboring point k_1 (in "k space") we have

$$\omega(k_1) = \omega(k_2) + (k_1-k_2)\left(\frac{d\omega}{dk}\right)_{k_2} + \tfrac{1}{2}(k_1-k_2)^2\left(\frac{d^2\omega}{dk^2}\right)_{k_2} + \cdots . \tag{40}$$

For k_1 close enough to k_2 we can neglect $(k_1 - k_2)^2$ and higher terms, and write

$$\boxed{v_{beat} = \frac{\omega_1 - \omega_2}{k_1 - k_2} \approx \left(\frac{d\omega}{dk}\right)_{k_{av}}} \tag{41}$$

where, since we are neglecting higher terms, we have evaluated $d\omega/dk$ at k_{av} instead of k_2. Thus we see that the "signal," consisting of the amplitude modulation $A(z,t)$, propagates not at the phase velocity but at the beat velocity.

For more general superpositions of traveling harmonic waves the same phenomenon occurs: If a number of waves of different frequencies in some band $\Delta\omega$ are superposed, the resultant traveling wave will be amplitude modulated. The amplitude modulations will propagate with a velocity called the group velocity, v_g, given by

$$v_g = \left(\frac{d\omega}{dk}\right)_{\omega_{av}} \tag{42}$$

where the derivative is evaluated at an average frequency ω_{av} in the band $\Delta\omega$.

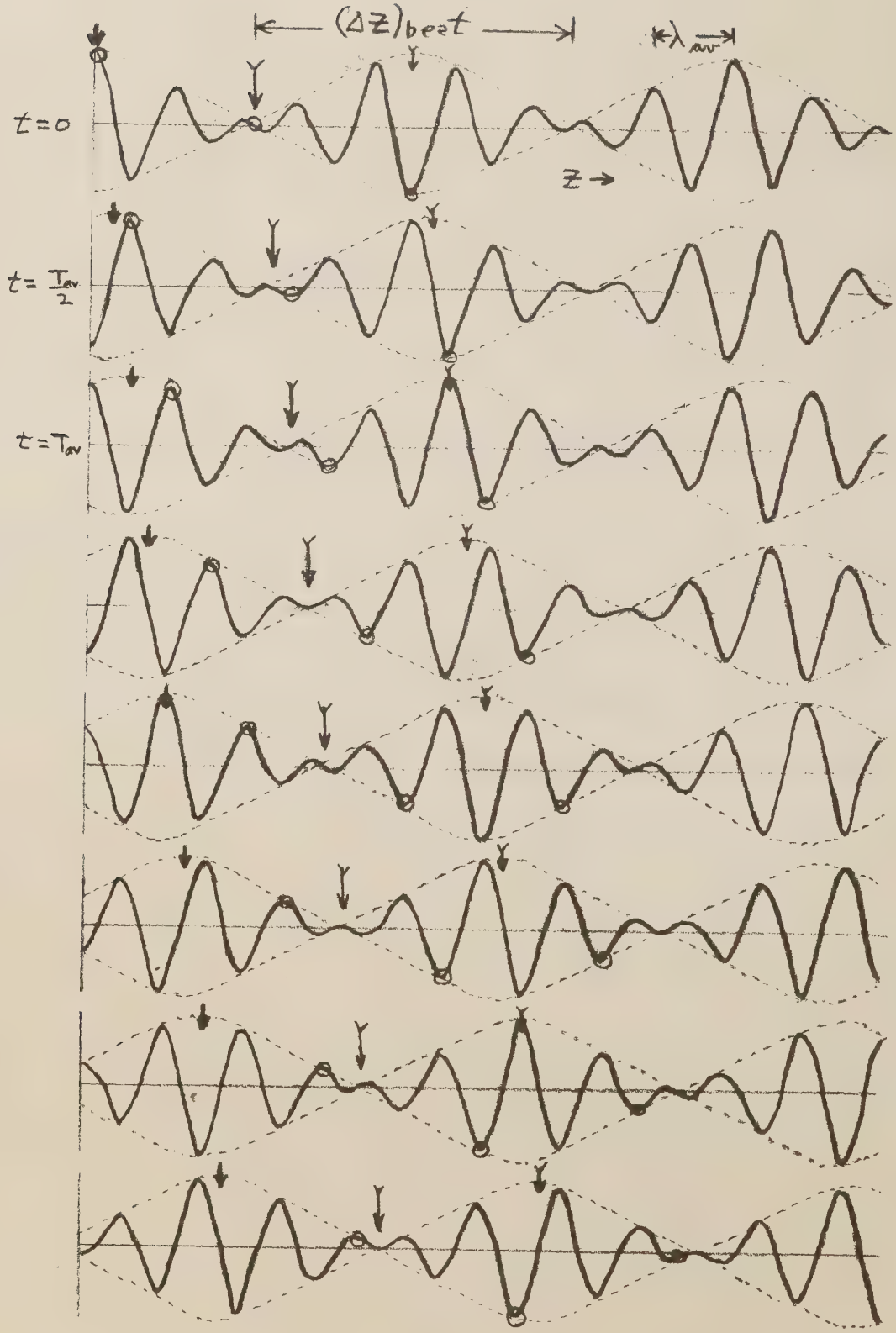


Fig. 2-2. Group velocity. The arrows follow the beats, traveling at the group velocity. The open circles follow individual wave crests traveling at the average phase velocity v_{av}.

In Fig. 2 we show the propagation of the "amplitude-modulated carrier wave" Eq. (25) for $\omega_{\text{modulation}} \equiv \omega_b = \omega_{av}/8$ and with the assumption that the phase velocity is twice the group velocity, as is the case for gravitational deep-water waves.

Here are some examples of group velocities.

Beaded string. The dispersion relation [Eq. (95), Chap. 1] for the beaded string,

$$\omega = 2(T/Ma)^{\frac{1}{2}} \sin \tfrac{1}{2} ka \quad , \tag{43}$$

gives a group velocity

$$v_g = \frac{d\omega}{dk} = a\,(T/Ma)^{\frac{1}{2}} \cos \tfrac{1}{2} ka = (T/\rho)^{\frac{1}{2}} \cos \tfrac{1}{2} ka \quad , \tag{44}$$

where $\rho = M/a$. The phase velocity is [Chap. 1, Eq. (93)]

$$v = \frac{\omega}{k} = (T/\rho)^{\frac{1}{2}} \frac{\sin \frac{1}{2} ka}{\frac{1}{2} ka} \quad . \tag{45}$$

Notice that in the limit $ka \to 0$, ω is a constant times k , and the waves are non-dispersive; the phase and group velocities are equal, each reducing to $(T/\rho)^{\frac{1}{2}}$

Note that if $ka = \pi$, the group velocity goes to zero. The beads are then one half wavelength apart,and there are not enough beads to tell which direction the waves are traveling. At this limit k_{max} the corresponding frequency is

$$\omega_{max} = 2(T/Ma)^{\frac{1}{2}} \quad . \tag{46}$$

If the transmitter at $z = 0$ shakes the string with frequency $\omega > \omega_{max}$, what happens? There are no traveling waves of the form given by Eq. (6). The "downstream" beads cannot "keep up" with their upstream neighbors except by undergoing smaller excursions. Thus the return force per unit

mass per unit displacement, which gives ω^2, can only be made large enough by diminishing the magnitude of the displacement of successive beads. The magnitude of the displacement turns out to be exponentially decreasing with distance, with the form of a "damped saw tooth,"

$$\psi(z_n,t) = A(-1)^n e^{-|k|z_n} \cos \omega t \quad . \tag{47}$$

The dispersion relation, instead of being Eq. (43), turns out to be

$$\omega = 2(T/Ma)^{\frac{1}{2}} \cosh \tfrac{1}{2} |k| a \quad . \tag{48}$$

We see here an example of a "cutoff" frequency, analogous to what we found in studying the dispersion relation for the ionosphere, (Chap. 1, Sec. 1.6.) where the plasma frequency ω_p is a cutoff frequency for radio waves. For the string, low frequencies can give traveling waves, high ones are "cut off." For the ionosphere it is just the opposite; the low frequencies are cut off and the high ones can give traveling waves.

Deep-water waves. (See Chap. 1, Sec. 1.6.) The dispersion relation is

$$\omega^2 = gk + \frac{Tk^3}{\rho} \quad . \tag{49}$$

The corresponding group velocity is

$$v_g = \tfrac{1}{2}\omega^{-1}\left[g + \frac{3Tk^2}{\rho}\right] \quad . \tag{50}$$

The phase velocity can be written in a form analogous to Eq. (50),

$$v = \omega^{-1}\left[g + \frac{Tk^2}{\rho}\right] \quad . \tag{51}$$

By comparison of Eqs. (50) and (51), we see that if we can neglect the

surface tension T, the group velocity is half the phase velocity. (We can neglect surface tension for waves for which $g \gg Tk^2/\rho$. This condition holds when the wavelength is large compared to about 1.7 cm, the wavelength for which g and Tk^2/ρ are equal.) On the other hand, for millimeter waves, where surface tension dominates, the group velocity is $1\frac{1}{2}$ times the phase velocity.

Electromagnetic waves in the ionosphere. This was discussed in Sec. 1.6, Chap. 1. It is important to recognize that the group velocity is always less than (or at most equal to) $c = 3.0 \times 10^{10}$ cm/sec. The group velocity $d\omega/dk$ is the velocity at which signals propagate in the form of amplitude modulations and is thus the velocity at which energy travels (the wave energy is maximum where the wave amplitude is maximum and is zero where the amplitude is zero). The phase velocity ω/k is the velocity at which individual wavelets travel. But individual wavelets do not "persist." If the phase velocity exceeds the group velocity v_g, the wavelets are "born" at the rear of a beat, travel to the front, and then die out at the front of the beat. If the phase velocity is less than the group velocity, they grow at the front of the beat, drop behind, and disappear at the rear of the beat. The beats or amplitude modulations (or, in the next section, the pulses or wave packets) carry the energy. Of course if v_g and v are equal, then the individual wavelets persist and maintain their position relative to the beat or wave group.

Esoteric examples. In Advanced Topic 3 we derive the phase and group velocities for de Broglie waves.

2.4 Superposition of many harmonic waves. Pulses and wave packets.

We wish to consider a situation where the transmitter at $z = 0$

describes a motion which is a superposition of many harmonic oscillations with equal amplitude and with closely-neighboring frequencies lying in a narrow band between the lowest, ω_1, and the highest, ω_2. We have already considered the situation where there are only two frequencies. In that case we get modulations (beats) which propagate down the string with the group velocity.

In preparation for the more complicated case involving many harmonic components with slightly differing frequencies, we reexamine the case of just two frequencies, using the technique of the rotating-vector diagram. (See Vol. 1, p. 125.) The harmonic oscillation

$$\psi(t) = A \cos \omega t \tag{52}$$

is the real part of the complex harmonic oscillation

$$\psi_c(t) = A e^{i\omega t} \tag{53}$$

where the subscript c stands for complex. A graphical representation of $\psi_c(t)$ is given by a vector of length A in the complex plane rotating counterclockwise at angular frequency ω. [The projection of this rotating vector on the horizontal axis (i.e., the real axis) gives the harmonic motion Eq. (52).] Instead of trying to visualize this rotating vector throughout one cycle we shall imagine that we take "stroboscopic snapshots." Then, if the stroboscope has the same frequency as the rotating vector, the vector will appear to stand still, i.e., every snapshot will catch the vector in the same position. See Fig. 3(a). If the angular frequency ω of the rotating vector is slightly greater than that of the stroboscope, ω_{strobe}, the vector will appear to rotate slowly counterclockwise at the

difference angular frequency $\omega-\omega_{strobe}$ [see Fig. 3(b)]; if, instead, $\omega-\omega_{strobe}$ is negative, the vector will appear to rotate slowly in the "retrograde" direction, i.e., clockwise [see Fig. 3(c)].

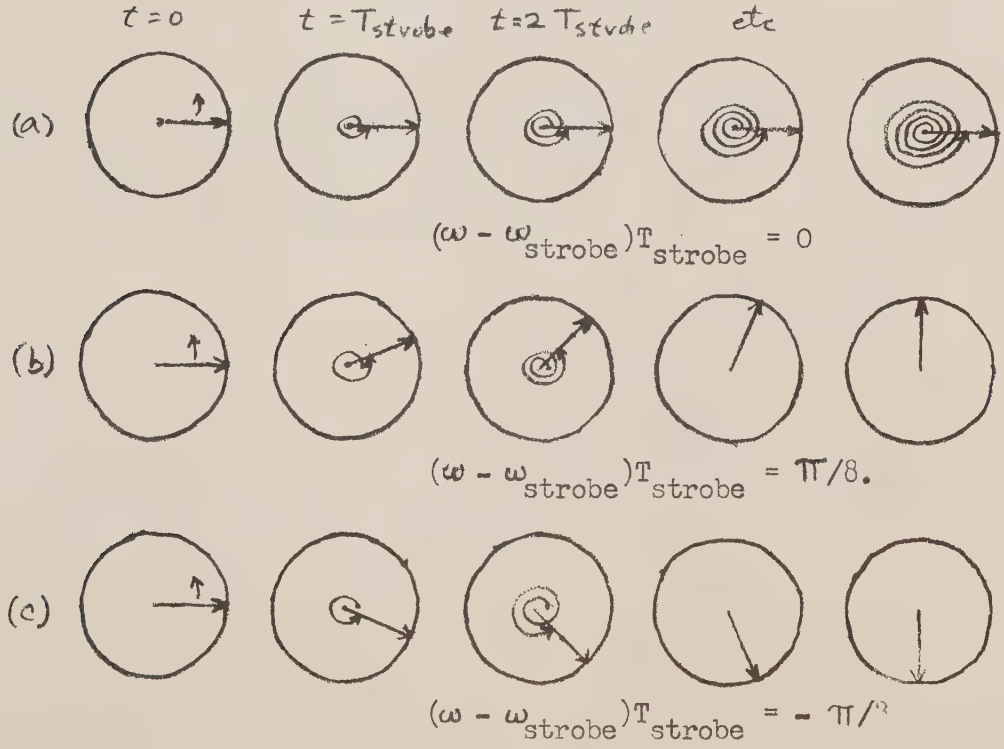


Fig. 2-3. Stroboscopic snapshots of rotating complex vector $e^{i\omega t}$. The spiral is to help you keep track of the number of rotations of the vector. The time interval between snapshots is $T_{strobe} = 2\pi/\omega_{strobe}$..

Now let us consider a superposition of two harmonic waves having the same amplitude but slightly different frequencies

$$\psi(t) = A \cos \omega_1 t + A \cos \omega_2 t \quad . \tag{54}$$

We "strobe" (take stroboscopic snapshots of) the rotating vectors $Ae^{i\omega_1 t}$ and $Ae^{i\omega_2 t}$ at the frequency

$$\omega_{strobe} = \omega_{av} = \tfrac{1}{2}(\omega_1 + \omega_2) \quad . \tag{55}$$

Thus (taking $\omega_2 - \omega_1$ to be positive) $\omega_2 - \omega_{av}$ is positive and $\omega_1 - \omega_{av}$ is negative. Recall that $\psi(t)$ may be written [as in Eq. (29)] as the product of a slowly varying amplitude $A(t)$ times a fast oscillation of frequency ω_{av}. Our strobe frequency of ω_{av} will cause the fast oscillation to "stand still," and only $A(t)$ will change between snapshots. Thus we obtain the snapshots shown in Fig. 4.

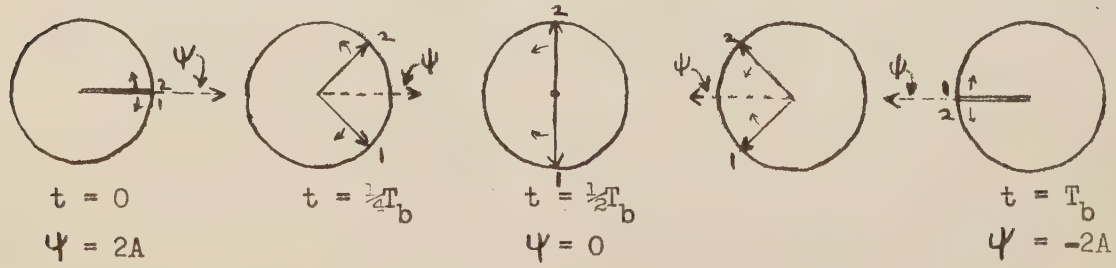


Fig. 2-4. Beats in the superposition $\psi(t) = Ae^{i\omega_1 t} + Ae^{i\omega_2 t}$. The snapshots are taken with $\omega_{strobe} = \omega_{av}$, and altogether cover exactly one beat of period T_b. (In this example the beat frequency is $\frac{1}{4}$ the average frequency; i.e., $\omega_2 - \omega_1 = \frac{1}{4}\omega_{av}$.) This figure represents graphically the content of Eqs. (22) or Eq. (29).

Now we consider the situation where $\psi(t)$ is a superposition of very many oscillations, all of equal amplitude A, phase constant zero, and distributed uniformly in frequency between ω_1 and ω_2. The corresponding stroboscopic vector diagram is shown in Fig. 5.

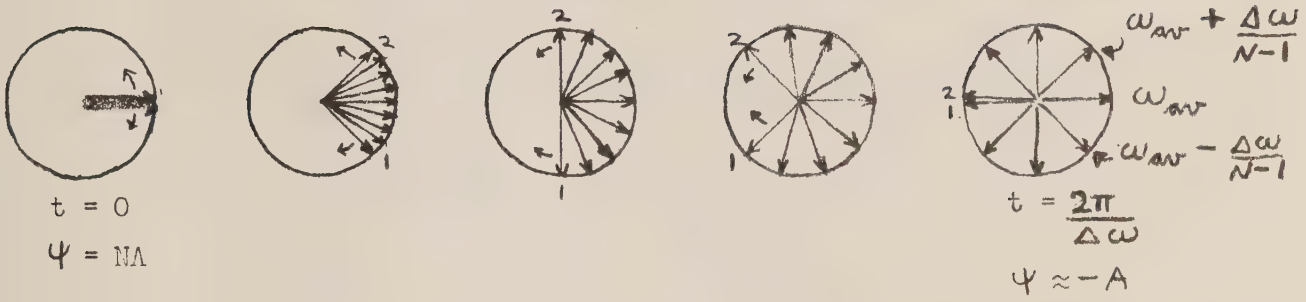


Fig. 2-5. N oscillations (N = 9 here) distributed uniformly within the frequency interval $\Delta\omega = \omega_2 - \omega_1$.

At $t = 0$ the total amplitude $A(t)$ of the superposition ψ is NA . At t slightly before $2\pi/\Delta\omega$, the period for a beat between ω_2 and ω_1 , the total amplitude $A(t)$ is zero, since the component contributions are uniformly distributed in phase. (When $N \to \infty$ this first zero occurs at exactly $t = 2\pi/\Delta\omega$.) For a long time after $t = 2\pi/\Delta\omega$, the vector contributions will still be widely distributed in phase, although not quite uniformly, so that the total amplitude $A(t)$ remains small for a long time. The vectors all reach the same phase again (and $A(t)$ returns to its original value of NA) only when the beats between contributions neighboring in frequency again reach their maxima. Since neighboring contributions have frequency separations $\Delta\omega/(N-1)$, the period for beats between neighboring frequencies is $(N - 1)$ times the beat period corresponding to $\Delta\omega$. Thus if $N \to \infty$, the total amplitude $A(t)$ remains small "forever," never returning to its original value. We then have what is called a pulse, i.e., a function of time that is significantly different from zero only during a limited time interval.

The exact solution for our particular example can be obtained with a

moderate amount of algebra. Thus if we have N contributions

$$\psi(t) = A \cos \omega_1 t + A \cos(\omega_1 + \delta\omega)t + A \cos(\omega_1 + 2\delta\omega)t + \ldots + A \cos \omega_2 t \qquad (56)$$

where $\delta\omega$ is the frequency spacing between neighboring contributions, i.e.,

$$\delta\omega \equiv \frac{\omega_2 - \omega_1}{N - 1} = \frac{\Delta\omega}{N - 1} , \qquad (57)$$

then one finds [see Appendix, Eqs. (21) to (34)]

$$\psi(t) = A(t) \cos \omega_{av} t , \qquad (58)$$

where

$$A(t) = \frac{A \sin\left[\frac{1}{2}\left(\frac{N}{N - 1}\right)\Delta\omega t\right]}{\sin\left[\frac{1}{2}\frac{\Delta\omega t}{N - 1}\right]} . \qquad (59)$$

For $N = 2$, Eq.(59) reduces to $A(t)$ in Eq. (29) and gives beats. For N so huge that we can take the argument of the denominator of Eq. (59) to be small compared with 1 radian for all times of interest (and so that $N - 1 \approx N$), we can rewrite Eq. (59) in the slightly simpler form

$$A(t) = NA \frac{\sin \frac{1}{2}\Delta\omega t}{\frac{1}{2}\Delta\omega t} = A(0) \frac{\sin \frac{1}{2}\Delta\omega t}{\frac{1}{2}\Delta\omega t} . \qquad (60)$$

In the second equality of Eq.(60) we got rid of the explicit N, with the understanding that as $N \to \infty$, A goes to zero in such a way that NA remains finite.

In Fig. 6 we plot Eq. (58) along with a "frequency spectrum" [to remind you that we started with Eq. (56)] corresponding to N contributions with equal amplitudes A and the same phase constant (zero), uniformly spaced in the frequency interval between ω_1 and ω_2 .

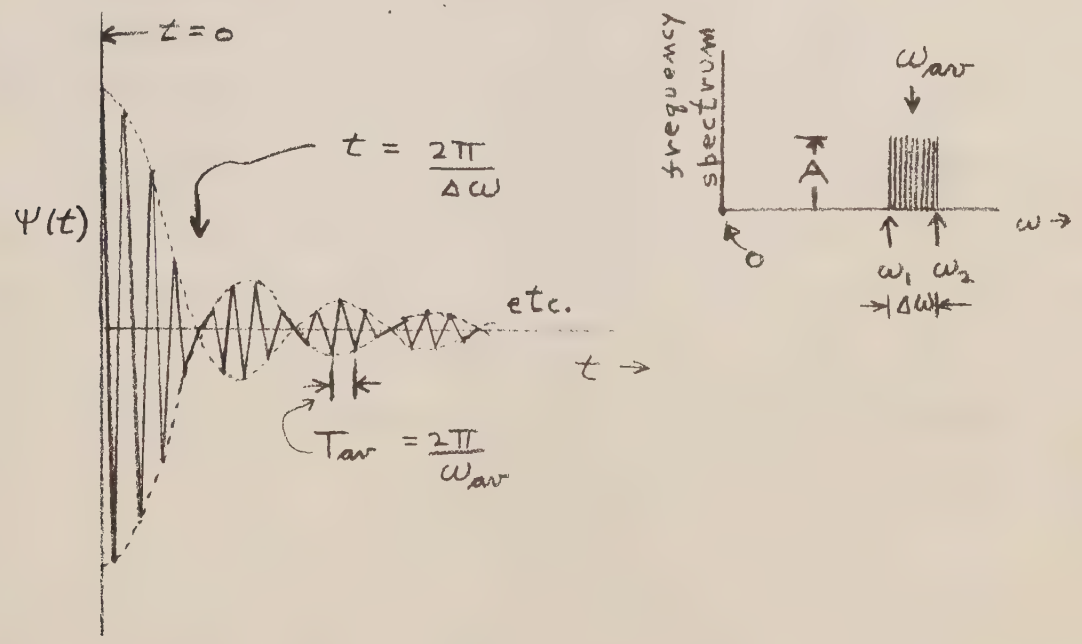


Fig. 2-6. A pulse $\psi(t)$ formed by superposing many frequency components in the narrow band of frequencies $\Delta\omega$. (For this example, $\omega_{av} = 4\,\Delta\omega$.) The frequency spectrum is shown in the insert; it corresponds to Eqs. (56) and (57). The superposition $\psi(t)$ corresponds to Eqs. (58) and (59).

Bandwidth-time interval product. From Fig. 6 we see that the pulse $\psi(t)$, made up of a superposition of harmonic components occupying a bandwidth $\Delta\omega$, has its first zero of amplitude at time $2\pi/\Delta\omega$ and, after that never amounts to much. Let us call the time interval from zero to $2\pi/\Delta\omega$, in which $\psi(t)$ is substantial, by the name Δt . Then we have $\Delta\omega\,\Delta t = 2\pi$,

i.e.,

$$\Delta\nu \, \Delta t = 1 \quad . \tag{61}$$

Equation (61) is a specific example of a very general (and very important) mathematical relationship between the time duration Δt of a pulse $\psi(t)$ and the bandwidth $\Delta\nu$ of the frequency spectrum of harmonic components which superpose to form the pulse. This relation is of broad applicability in all of physics, wherever there are phenomena with the form of a pulse in time or some other variable. The general relationship is independent of the detailed specific shape of $\psi(t)$, as long as $\psi(t)$ has the characteristic that defines a pulse, namely that $\psi(t)$ is substantially different from zero only during a single limited time interval of duration Δt. (In any specific instance one can give a precise definition of Δt, just as we did for the above example.) The general relationship is

$$\boxed{\Delta\nu \, \Delta t \approx 1} \quad , \tag{62}$$

where the "approximately equals" sign can be made an equals sign in any specific instance by suitable definitions of $\Delta\nu$ and Δt, as in Eq. (61).

Sometimes one expresses Eq. (62) by saying that there is an "uncertainty" in "the" frequency. Thus, looking at Fig. 6, we would say that $\psi(t)$ is dominated by oscillation of frequency ω_{av}, but, since there is actually a small band of frequencies occupying a band of width $\Delta\omega$, we cannot say that "the" frequency is ω_{av} -- there is an "uncertainty" as to what the frequency is, and the extent of this uncertainty is the bandwidth $\Delta\omega$. Any frequency in this band is about as good a candidate for the title of "the" frequency as any other.

Here are some further illustrations of Eq. (62).

Harmonic oscillation of limited duration. Suppose an oscillator is turned on, builds up rapidly (in a few cycles) to constant amplitude A, oscillates for a number of cycles according to the harmonic oscillation $A \cos \omega_o t$, is then turned off and dies out in a few cycles, as shown in Fig. 7.

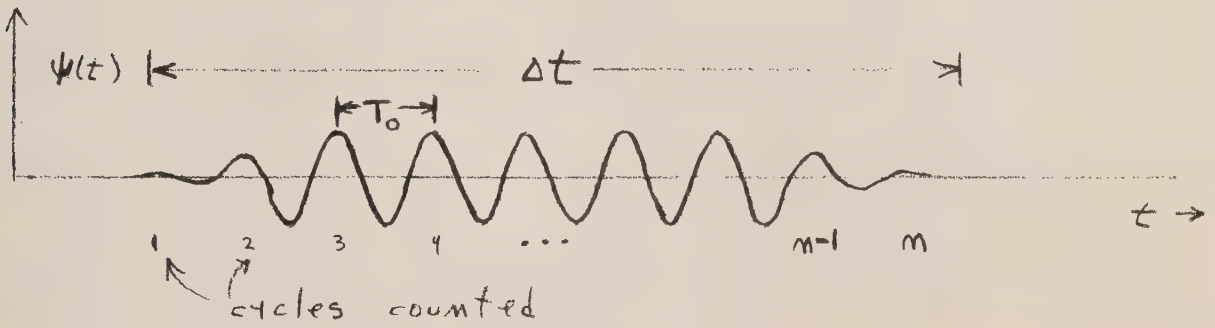


Fig. 2-7. Harmonic oscillation of limited duration.

What is "the" frequency? Certainly the (angular) frequency $\omega = \omega_o$ dominates. But, according to what we have said, it cannot be the only frequency ω present. Let us use the "uncertainty" approach. We shall try to find "the" frequency, and then ask, "what is the uncertainty in our estimate?" Using the definition of frequency as cycles per second, we simply count the total number of cycles in the time interval Δt during which the oscillator is turned on and divide by Δt. Thus, counting n cycles, we find

$$\nu \text{ (dominant)} = \frac{n}{\Delta t} , \tag{63}$$

which is bound to be about equal to $\nu_o = T_o^{-1}$, according to Fig. 7. Now for the uncertainty. We see by inspection of Fig. 7 that it is not possible to specify n exactly. There is an uncertainty of order $\pm\frac{1}{2}$ cycle at each end of the pulse, where we are having to decide "shall we count one more, or

is it dead?" You may remark, "It doesn't matter very much, especially if n is large -- the error is small compared with n." Yes, but it is exactly that error that we are concerned with. According to Eq. (63), an uncertainty band Δn of width approximately 1 in the number of cycles n gives an uncertainty band $\Delta\nu$ of width $\Delta n/\Delta t$, which is then approximately $1/\Delta t$, i.e.,

$$\Delta\nu = \frac{\Delta n}{\Delta t} \approx \frac{1}{\Delta t} \quad . \tag{64}$$

This is the same relation as Eq. (62)

When one uses the uncertainty point of view, one applies it to the time also. Thus, suppose someone says, "Exactly when, i.e., at exactly which instant was the oscillator on?" Of course there is no such single instant. It was on all during Δt. But if he insists on an answer, you can choose the center of the interval Δt and say that that is when it was on, with, however, an uncertainty of width Δt to be attached to your answer.

It is better for us if we do not use the uncertainty language. It is more sensible to simply say that an oscillation of dominant frequency ν_o that lasts for a time interval Δt contains a spectrum of frequencies centered at ν_o and of band width $\Delta\nu$, satisfying Eq. (62).

*Weakly-damped harmonic oscillator. Consider a damped harmonic oscillator for which $\psi(t)$ is zero from $t = -\infty$ until $t = 0$, at which time the oscillator is suddenly given a burst of energy and then allowed to oscillate. Then the oscillation is given by

$$\psi(t) = A\, e^{-\frac{1}{2}t/\tau} \cos \omega_o t \quad . \tag{65}$$

The factor of $\frac{1}{2}$ in the exponential-decay factor is put there so that the

energy $E(t)$ (potential plus kinetic energy) of the oscillator will decay with mean decay time τ, i.e., (see Volume 1, p. 210)

$$E(t) = E(0)e^{-t/\tau} . \tag{66}$$

We may express the frequency spectrum as a sum over contributions of the form $A(\omega) \cos [\omega t + \phi(\omega)]$. (In our earlier example of a frequency spectrum, Eq. (56), $A(\omega)$ is the constant A and $\delta(\omega)$ equals zero for all ω in the band $\Delta\omega$. That is about the simplest possible frequency spectrum.) As you may learn in a future course that treats Fourier integrals, it turns out that the frequency spectrum of Eq. (65) is given approximately by

$$A^2(\omega) = \frac{\text{constant}}{(\omega - \omega_o)^2 + (\frac{1}{2}\tau^{-1})^2} \tag{67}$$

and

$$\tan \delta(\omega) = 2\tau(\omega_o - \omega) . \tag{68}$$

In Fig. 8 we plot $\psi(t)$ and $A^2(\omega)$. The "intensity" component $A^2(\omega)$

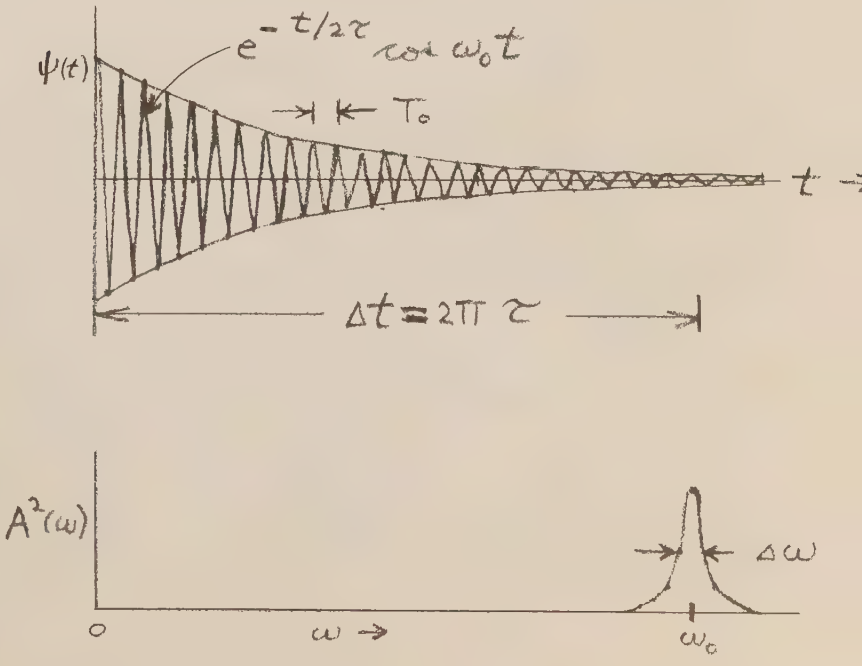


Fig. 2-8. Damped harmonic oscillator. The frequency spectrum has a full-width at half maximum of $\Delta\omega$, so that $\Delta\omega\,\tau = 1 = \Delta\nu\,\Delta t$, with $\Delta t = 2\pi\tau$. For this example, $\tau = 4T_o$, so that $\Delta\omega/\omega_o = 1/8\pi$.

(called intensity because it is the square of an amplitude), as given by Eq. (67), decreases by a factor $\frac{1}{2}$ when $\omega - \omega_o$ goes from zero to $\pm \frac{1}{2} \tau^{-1}$, because the denominator then doubles. Calling the range of ω between these two limits by the name $\Delta\omega$ [which makes $\Delta\omega$ the "full width at half maximum" for the plot of $A^2(\omega)$ versus ω], we see that we have

$$\Delta\omega = \omega_{max} - \omega_{min} = (\omega_o + \tfrac{1}{2}\tau^{-1}) - (\omega_o - \tfrac{1}{2}\tau^{-1}) = \tau^{-1} ,$$

i.e.,

$$\boxed{\Delta\omega\,\tau = 1} \qquad (69)$$

Equation (69) is exact, because we have given precise definitions of $\Delta\omega$ and τ. To get agreement with the general result Eq. (62), we can take for Δt, a measure of the duration of the decaying oscillation $\psi(t)$ [Eq. 65)],

$$\Delta t = 2\pi\,\tau . \qquad (70)$$

Inserting this into Eqs. (69) yields

$$\Delta\nu\,\Delta t = 1 , \qquad (71)$$

where we have an exact equality instead of the approximate equality of Eq. (62) because we have specified $\Delta\nu$ and Δt precisely.

Amplitude-modulated radio waves. Suppose an orchestra is playing into a microphone. The sound waves have frequencies from 20 cps to about 20 Kc. (Those are roughly the limits of human hearing.) The voltages induced in the microphone at sound frequencies ν_m are used to modulate the amplitude of a radio-frequency "carrier wave." Typical carrier wave frequencies ν_o

for AM radio are 500 to 1500 Kc. The current in the transmitting antenna is of the form

$$I(t) = I_o \cos \omega_m t \cos \omega_o t = \tfrac{1}{2} I_o \cos(\omega_o+\omega_m)t + \tfrac{1}{2} I_o \cos(\omega_o-\omega_m)t \quad , \qquad (72)$$

for a given modulation angular frequency ω_m (the sound frequency). In order to carry all the "information" played by the orchestra one needs a "lower side band" $\nu_o - \nu_m$ from ν_o to ν_o - 20 Kc , and an "upper side band" $\nu_o + \nu_m$ from ν_o to ν_o + 20 Kc . Thus to transmit music without distortion requires a bandwidth $\Delta\nu \approx 40$ Kc . Between 500 Kc and 1500 Kc there is therefore "air space" (bandwidth) available for about 1000/40 = 25 AM stations, without overlap. (Actually, one of the sidebands is ingeniously suppressed -- filtered out and never applied to the transmitting antenna. It is then regenerated in the receiver. This halves the bandwidth requirement. This is called "single side-band" broadcasting.)

Television. The picture on the TV screen consists of a rectangular grid pattern of black and white spots. A given spot is "white" if the phosphorescent TV screen was recently (within $\sim$ 1/50 sec) struck by the electron beam at that spot. The spot separation is about 1 mm. A typical screen 50 cm × 50 cm thus has 500 lines with 500 spots per line, or 25×10^4 spots. The pictures are renewed every 1/30 sec. Therefore the rate at which the instructions "turn off, turn on, turn on, ..." must be sent to the electron beam is about $30 \times 25 \times 10^4$ or 8×10^6 times per second. The transmitted and received antenna voltages must therefore have about 10^7 little voltage on-off bumps per second. Each bump can last no longer than $\Delta t \sim 10^{-7}$ sec, to prevent overlap. Therefore the required bandwidth is $\Delta\nu \approx 1/\Delta t \approx 10^7$ cps = 10 Mc. Thus between $\nu_o \approx 55$ Mc and 210 Mc there is only enough bandwidth

for about 12 TV channels, assuming an average separation of about 13 Mc.

Traveling wave packet. Suppose the transmitter at $z = 0$ describes a motion in the form of a pulse similar to that of Fig. 7. The corresponding traveling waves that propagate in the medium form what is called a "wave packet" or "wave group." The wave packet propagates with the group velocity. (Any one of the beats in Fig. 2 can be thought of as a wave packet.) Because k and ω are related through the dispersion relation $k(\omega)$, the existence of a band $\Delta\omega$ of frequencies emitted by the transmitter implies a corresponding band Δk of wave numbers (and corresponding wavelengths) in the wave packet. Related to the dominant frequency ω_o there will be a dominant wave number $k_o = k(\omega_o)$ [i.e., k_o is obtained by substituting $\omega = \omega_o$ in the functional relation $k(\omega)$]. The band Δk is centered at k_o and is obtained by differentiating the dispersion relation and setting $\omega = \omega_o$.

$$\Delta k = \left(\frac{dk}{d\omega}\right)_o \Delta\omega = \frac{\Delta\omega}{v_g} , \tag{73}$$

where we used $v_g = (d\omega/dk)_o$. (The subscript zero means that the derivative is evaluated at the center of the band. Also, we neglect higher terms in the Taylor's series expansion of the dispersion relation for which Eq. (73) may be considered the first term.)

A packet of length Δz traveling at group velocity v_g passes a given fixed point z in a time interval Δt given by

$$\Delta z = v_g \Delta t . \tag{74}$$

Taking the product of Eqs. (73) and (74), we obtain

$$\Delta k \, \Delta z \;=\; \Delta \omega \, \Delta t \quad . \tag{75}$$

Thus from Eq. (62) we have

$$\Delta \, (k/2\pi) \, \Delta z \;\approx\; 1 \quad ,$$

i.e.,

$$\boxed{\Delta \, (\lambda^{-1}) \, \Delta z \;\approx\; 1} \quad . \tag{76}$$

This relation is really just another case of the general relation Eq. (62), but with a pulse in space instead of time.

Another simple way to derive Eq. (76) is to consider the uncertainty in the number of cycles contained in Δz . Thus

$$\lambda^{-1} \;=\; \frac{\text{cycles} \pm \frac{1}{2}}{\Delta z} \quad , \tag{77}$$

so that the uncertainty $\Delta(\lambda^{-1})$ in λ^{-1} is approximately $1/\Delta z$. This derivation is the analogue in space of the derivation of Eq. (64) in time.

Finally we should point out that the length Δz of a packet does not remain constant as the packet propagates in the medium. The packet spreads. That is because there is variation of the group velocity $v_g = d\omega/dk$ with k (or ω). Therefore the band Δk contains a band of group velocities Δv_g given approximately by

$$\Delta v_g \;=\; \left(\frac{dv_g}{dk}\right)_o \Delta k \;=\; \left(\frac{d^2\omega}{dk^2}\right)_o \Delta k \quad , \tag{78}$$

A group which starts out with width $(\Delta z)_o$ at $t = 0$ will have a width $(\Delta z)_t$ at time t given approximately by

$$(\Delta z)_t \;\approx\; (\Delta z)_o + (\Delta v_g)\, t \quad . \tag{79}$$

The time Δt that the packet takes to pass a fixed z will increase accordingly. [Equation (75) holds at all times and, of course, $\Delta\omega$ and Δk are constant.]

Because of this spreading of the packet, the relations $\Delta(\lambda^{-1})\,\Delta z \approx 1$, and $\Delta\nu\,\Delta t \approx 1$ do not hold except at $t = 0$. That seems mysterious. Didn't we prove $\Delta\nu\,\Delta t \approx 1$? Yes, but notice that in exercising our freedom to build up a superposition of harmonic components to represent the pulse $\psi(t)$, we took all harmonic components to be in the correct phase at $t = 0$. However, once we let the group propagate down the medium, we cannot have the entire band Δk in phase at a downstream point, because some parts of the group arrive a little early and some late due to the variation in group velocity. Thus the phases of all parts of the wave are all different from one another at a downstream point, unlike the situation at $t = 0$.

Wave packets in water. You can make a nice expanding circular wave packet by tossing a pebble in a pond. With some practice you can follow a packet with your eyes and watch individual wavelets grow at the back, pass through the packet, and "disappear" at the front. [The phase velocity is greater than the group velocity for $\lambda > 1.7$ cm, as is usually the case for a moderate-sized pebble. See Eqs. (50) and (51).] A picture of a wave group with phase velocity twice the group velocity is shown in Fig. 9.

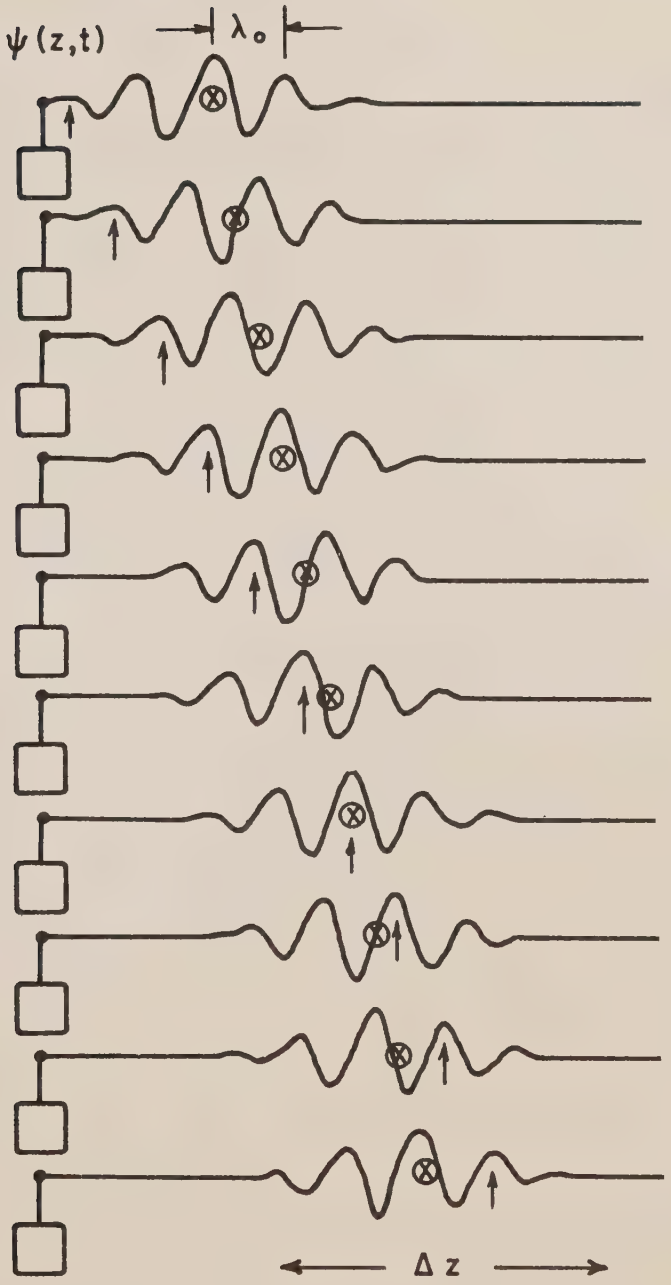


Fig. 2-9. Wave packet with phase velocity twice the group velocity. The arrow travels at the phase velocity, following a point of constant phase for the dominant wavelength. The cross travels at the group velocity with the packet as a whole.

2.5 Wave equations.

We have been working directly with waves and the continuous physical phenomena which result in wave motion--either standing waves or traveling waves, depending on the boundary conditions. One very general way to express important properties of the waves for any given physical situation is to give the dispersion relation which relates ω and k. (See Sec. 1.6.) Another equivalent way is to give the wave equation, which is a linear partial differential equation (for waves satisfying the superposition principle) giving the relation between $\psi(z,t)$, $\partial\psi(z,t)/\partial t$, $\partial^2\psi(z,t)/\partial t^2, \cdots$, $\partial\psi(z,t)/\partial z$, $\partial^2\psi(z,t)/\partial z^2, \cdots$ etc. One can learn about the behavior of the waves by studying the solutions of the wave equation, with appropriate boundary conditions. That is usually too "mathematical" an approach for us. However, it is a powerful approach. Sometimes phenomena of entirely different physical natures have the "same" wave equation. Therefore if you are familiar with the solutions of the wave equation for one physical phenomenon you may be able to understand another entirely different phenomenon by analogy. (It sometimes doesn't work, usually because the boundary conditions are not analogous.)

We shall start with harmonic waves and find linear relations between the wave functions $\psi(z,t)$ and their derivatives at a given point (z,t). Then we shall use the dispersion relation to obtain the wave equation. We shall consider standing waves and traveling waves at the same time. For the following five equations, the left half of the page will give the standing wave formulas, and the right half will give these for traveling waves (for waves traveling in the $+z$ direction).

Standing waves	Traveling waves	
$\psi(z,t) = A \cos kz \cos \omega t$	$\psi(z,t) = A \cos(kz - \omega t)$	(80)
$\frac{\partial \psi(z,t)}{\partial t} = -\omega A \cos kz \sin \omega t$	$\frac{\partial \psi(z,t)}{\partial t} = \omega A \sin(kz - \omega t)$	(81)
$\frac{\partial^2 \psi(z,t)}{\partial t^2} = -\omega^2 \psi(z,t)$	$\frac{\partial^2 \psi(z,t)}{\partial t^2} = -\omega^2 \psi(z,t)$	(82)
$\frac{\partial \psi(z,t)}{\partial z} = -kA \sin kz \cos \omega t$	$\frac{\partial \psi(z,t)}{\partial z} = -kA \sin(kz - \omega t)$	(83)
$\frac{\partial^2 \psi(z,t)}{\partial z^2} = -k^2 \psi(z,t)$	$\frac{\partial^2 \psi(z,t)}{\partial z^2} = -k^2 \psi(z,t)$	(84)

We see that the second derivatives with respect to t and to z give results which are the same for standing waves as for traveling waves [i.e., Eqs. (82) and (84)]. The wave equation can be found if the dispersion relation is known. Here are some examples:

Non-dispersive waves. The Classical Wave Equation. Let us take the phase velocity $v = \omega/k$ to be independent of ω and k. Substituting v^2k^2 for ω^2 in Eq. (82) (for either standing or traveling waves) and combining the result with Eq. (84) gives

$$\frac{\partial^2 \psi(z,t)}{\partial t^2} = v^2 \frac{\partial^2 \psi(z,t)}{\partial z^2} . \qquad (85)$$

Equation (85) is often called "the classical wave equation." It holds whenever the dispersion relation is such that $v = \omega/k$ is a constant independent of ω and k. Its most general solution can be written as a superposition of standing harmonic waves, or alternatively, as a superposition of traveling harmonic waves (traveling in both +z and -z directions). It applies to ordinary sound waves with great accuracy. It also applies to light in vacuum, with $v = c = 3 \times 10^{10}$ cm/se

(It does not apply to light in glass or water.) It applies to the stretched beaded string in the limit that ka may be neglected compared to 1, i.e., for $a \to 0$, or for wavelength large compared to a.

Notice that for a single harmonic component of frequency ω Eqs. (82) and (84) give

$$\frac{\partial^2 \psi(z,t)}{\partial t^2} = v^2(k) \frac{\partial^2 \psi(z,t)}{\partial z^2} \tag{86}$$

If we write Eq. (86) twice, once for $k = k_1$ and once for $k = k_2$, and then add the two equations, we get

$$\frac{\partial^2 [\psi_1(z,t) + \psi_2(z,t)]}{\partial t^2} = v^2(k_1) \frac{\partial^2 \psi_1(z,t)}{\partial z^2} + v^2(k_2) \frac{\partial^2 \psi_2(z,t)}{\partial z^2} . \tag{87}$$

This can only be put in the form of Eq. (85) for the superposition $\psi_1 + \psi_2$ if $v^2(k_1)$ and $v^2(k_2)$ are equal, so that v^2 may be factored out in the right side of Eq. (87). Thus for dispersive waves (those where v(k) varies with k), Eq. (85) does not hold if more than one frequency is present.

Here is an interesting and useful way to express the solutions of Eq. (85). We know that in a non-dispersive medium all harmonic traveling waves travel at the same phase velocity v. Therefore for waves traveling in the +z direction (denoted by the subscript +),

$$\psi_+(z,t) = A_1 \cos(k_1 z - \omega_1 t + \alpha_1) + A_2 \cos(k_2 z - \omega_2 t + \alpha_2) + \cdots$$

$$= A_1 \cos[k_1(z - vt) + \alpha_1] + A_2 \cos[k_2(z - vt) + \alpha_2] + \cdots \tag{88}$$

$$\equiv F(z - vt) . \tag{89}$$

It can be shown that a very general class of functions of the variable z - vt can be written in the form Eq. (88), which is called a Fourier series. Thus, for non-dispersive waves we can, if we wish, forego expressing $\psi(z,t)$ explicitly

in terms of harmonic traveling waves and simply say that $\psi_+(z,t)$ is <u>any</u> function $F_+(z - vt)$ of the argument $z - vt$.

Similarly, any wave function $\psi_-(z,t)$ that consists of a general superposition of traveling waves traveling in the -z direction can be written as a function of $z + vt$ (with v taken to be positive). Thus the most general solution of Eq. (85) can be written either as a superposition of harmonic waves traveling in both directions or as a superposition of two "arbitrary" traveling-wave functions $F(z - vt)$ and $G(z + vt)$, i.e.,

$$\psi(z,t) = F(z - vt) + G(z + vt). \tag{90}$$

We can check that $\psi(z,t)$ as given by Eq. (90) satisfies Eq. (85) for arbitrary functions F and G (as long as they are not so arbitrary that they cannot be differentiated). Thus, differentiating Eq. (90) gives

$$\frac{\partial\psi}{\partial t} = \frac{\partial F}{\partial t} + \frac{\partial G}{\partial t} = \frac{dF}{d(z - vt)}\,\frac{\partial(z - vt)}{\partial t} + \frac{dG}{d(z + vt)}\,\frac{\partial(z + vt)}{\partial t}$$

$$= -v\,\frac{dF}{d(z - vt)} + v\,\frac{dG}{d(z + vt)}\ . \tag{91}$$

$$\frac{\partial^2\psi}{\partial t^2} = -v\,\frac{d^2F}{d(z - vt)^2}\,\frac{\partial(z - vt)}{\partial t} + v\,\frac{d^2G}{d(z + vt)^2}\,\frac{\partial(z + vt)}{\partial t}$$

$$= v^2\left[\frac{d^2F}{d(z - vt)^2} + \frac{d^2G}{d(z + vt)^2}\right]\ . \tag{92}$$

Similarly,

$$\frac{\partial\psi}{\partial z} = \frac{\partial F}{\partial z} + \frac{\partial G}{\partial z} = \frac{dF}{d(z - vt)}\,\frac{\partial(z - vt)}{\partial z} + \frac{dG}{d(z + vt)}\,\frac{\partial(z + vt)}{\partial z}$$

$$= \frac{dF}{d(z - vt)} + \frac{dG}{d(z + vt)}\ . \tag{93}$$

$$\frac{\partial^2\psi}{\partial z^2} = \frac{d^2F}{d(z - vt)^2}\,\frac{\partial(z - vt)}{\partial z} + \frac{d^2G}{d(z + vt)^2}\,\frac{\partial(z + vt)}{\partial z}$$

$$= \frac{d^2F}{d(z - vt)^2} + \frac{d^2G}{d(z + vt)^2} \quad . \tag{94}$$

Combining Eqs. (92) and (94), we see that $\psi(z,t)$ as given by Eq. (90) satisfies the classical wave equation, Eq. (85).

The following is an example of a traveling wave in a non-dispersive medium with a form $F(z - vt)$ which is not that of a harmonic traveling wave (but could, of course, be written as a superposition of harmonic traveling waves):

*Gaussian pulse and wave packet in non-dispersive medium. Suppose that a transmitter at $z = 0$ describes a motion in the form of a pulse of displacement $D(t)$ given by

$$D(t) = A\, e^{-at^2} \equiv A \exp(-at^2) \quad , \tag{95}$$

where A and a are positive constants. This pulse has its maximum value, A, at $t = 0$. From $t = -\infty$ to $t = 0$ it builds up from nothing (at $t = -\infty$) to A (at $t = 0$). For times after $t = 0$ it dies away. Suppose the boundary condition at $z = 0$ is given by

$$\psi(0,t) = D(t) \quad , \tag{96}$$

as is the case for a string attached to the output hook of a transmitter. Then we can write down a function of $z - vt$ that satisfies the boundary condition Eq. (96) by simply replacing t by $t - (z/v)$ in $D(t)$:

$$\psi(z,t) = \psi[0,t - (z/v)] = D(t - z/v) \tag{97}$$

$$= A \exp\,[-a(t - z/v)^2]\,, \tag{98}$$

$$= A \exp\,[-(a/v^2)(vt - z)^2]\,, \tag{99}$$

which is an example of a function $F(z - vt)$.

The first equality in Eq. (97) is very general. It applies to all non-dispersive traveling waves (traveling in the +z direction). It simply says that ψ at place z at time t is the same as ψ at $z = 0$ at the earlier time $t - z/v$. This is true because the wave is traveling without changing its shape. It looks the same later, downstream, as it did earlier, upstream. That is <u>not</u> the case when the medium is dispersive, i.e., when $v(k)$ depends on k. Then we do not have $\psi(z,t)$ given by $F(z - vt) + G(z + vt)$, Eq. (90), and $\psi(z,t)$ does not satisfy the classical wave equation, Eq. (85). Of course a <u>single</u> harmonic traveling wave (i.e., with only one frequency present) does travel without changing its shape, but that is only a very special case.

The second equality in Eq. (97) is not general. It is a consequence of the particular assumption we made in Eq. (96) for the relationship between $\psi(0,t)$ and the motion $D(t)$ of the transmitter. (See problem 2 for an example of a different boundary condition.)

In Fig. 10 we show a pulse emitted at $z = 0$ and traveling down a non-dispersive string.

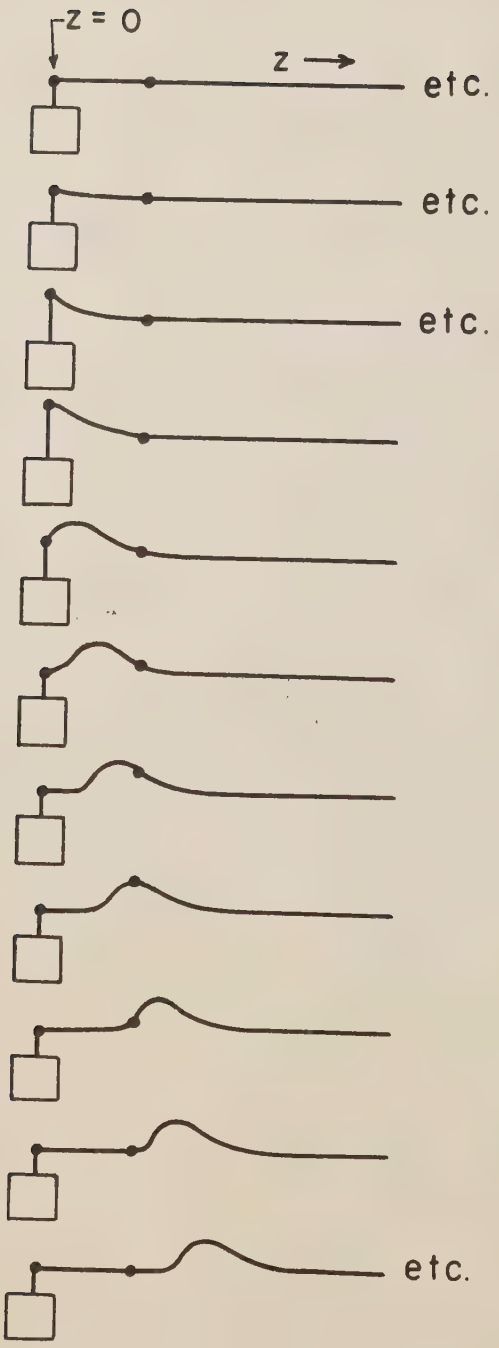


Fig. 2-10. Wave of transverse displacement traveling away from a disturbance at $z = 0$. Each point on the string undergoes the same motion as that of the point at $z = 0$, but at a later time.

Other wave equations.

Here are some examples, to satisfy your curiosity. We shall not actually work with these wave equations.

*Beaded string. The dispersion relation [Eq. (95), Sec 1.6)] is

$$\omega^2 = \frac{4T}{Ma}\sin^2(\tfrac{1}{2}ka) = \frac{T}{\rho}k^2 - \frac{1}{12}\frac{T}{\rho}a^2k^4 + \cdots , \tag{100}$$

where we have expanded $\sin^2(\frac{1}{2}ka)$ in a Taylor's series around $\frac{1}{2}ka = 0$. Substituting v_o^2 (the phase velocity for $ka \to 0$) for $\frac{T}{\rho}$, and using the relation $k^2\psi(z,t) = -\partial^2\psi(z,t)/\partial z^2$ [Eq. (84)] and the relation $k^4\psi(z,t) = \partial^4\psi(z,t)/\partial z^4$ (obtained by differentiating Eq. (84) twice more with respect to z), we can multiply Eq. (100) by $-\psi$ and, with the help of Eq. (82), finally obtain

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = v_o^2\frac{\partial^2\psi(z,t)}{\partial z^2} + \frac{1}{12}v_o^2a^2\frac{\partial^4\psi(z,t)}{\partial z^4} + \cdots . \tag{101}$$

If we drop all but the first term on the right-hand side of this horrible looking equation, we obtain the classical wave equation. Since we know the exact dispersion relation [the first equality in Eq. (100)] there is no particular reason for studying Eq. (101). Thank goodness.

*Electromagnetic waves in the ionosphere. The dispersion relation is (see Eq. (110), Sec. 1.6)

$$\omega^2 = c^2k^2 + \omega_p^2 . \tag{102}$$

The corresponding wave equation is obtained by multiplying Eq. (102) by $-\psi$ and using Eqs. (82) and (84):

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = c^2\frac{\partial^2\psi(z,t)}{\partial z^2} - \omega_p^2\psi(z,t). \tag{103}$$

In vacuum, ω_p is zero. Then we get the classical wave equation with $v = c$.

Electromagnetic waves in vacuum. We wish to show that Maxwell's equations for vacuum lead to the classical wave equation, which for three dimensional waves is given by

$$\frac{\partial^2 \psi(x,y,z,t)}{\partial t^2} = c^2 \nabla^2 \psi(x,y,z,t) , \tag{104}$$

where

$$c^2 \nabla^2 \psi \equiv c^2 \underset{\sim}{\nabla} \cdot \underset{\sim}{\nabla} \psi \equiv c^2 \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) . \tag{105}$$

Maxwell's equations for vacuum are given by (see Vol. 2, page 264)

$$\frac{\partial \underset{\sim}{E}}{\partial t} = c \underset{\sim}{\nabla} \times \underset{\sim}{B} \tag{106}$$

$$\frac{\partial \underset{\sim}{B}}{\partial t} = -c \underset{\sim}{\nabla} \times \underset{\sim}{E} \tag{107}$$

$$\underset{\sim}{\nabla} \cdot \underset{\sim}{E} = 0 \tag{108}$$

$$\underset{\sim}{\nabla} \cdot \underset{\sim}{B} = 0. \tag{109}$$

We shall find a wave equation for $\underset{\sim}{E}$. (A completely analogous derivation can be given for $\underset{\sim}{B}$.) We start by differentiating Eq. (106) with respect to t; on the right-hand side, we interchange the order of the time derivative and the curl; then we substitute $\partial \underset{\sim}{B}/\partial t$ from Eq. (107). Thus

$$\frac{\partial^2 \underset{\sim}{E}}{\partial t^2} = c \frac{\partial (\underset{\sim}{\nabla} \times \underset{\sim}{B})}{\partial t} = c \underset{\sim}{\nabla} \times \frac{\partial \underset{\sim}{B}}{\partial t} = -c^2 \underset{\sim}{\nabla} \times (\underset{\sim}{\nabla} \times \underset{\sim}{E}) . \tag{110}$$

It can be shown [Appendix, Eq. (41)] that for any vector $\underset{\sim}{C}$

$$\underset{\sim}{\nabla} \times (\underset{\sim}{\nabla} \times \underset{\sim}{C}) = \underset{\sim}{\nabla} (\underset{\sim}{\nabla} \cdot \underset{\sim}{C}) - (\underset{\sim}{\nabla} \cdot \underset{\sim}{\nabla}) \underset{\sim}{C} . \tag{111}$$

Substituting $\underset{\sim}{E}$ for $\underset{\sim}{C}$ in Eq. (111), and using the fact that $\underset{\sim}{\nabla}\cdot\underset{\sim}{E} = 0$ [Eq. (108)], we obtain from Eq. (110)

$$\frac{\partial^2 \underset{\sim}{E}(x,y,z,t)}{\partial t^2} = c^2 \nabla^2 \underset{\sim}{E}(x,y,z,t) \ . \tag{112}$$

This vector equation consists of three separate partial differential equations, each of the form of Eq. (104), with $\psi(x,y,z,t)$ equal to $E_x(x,y,z,t)$ in the first equation, $E_y(x,y,z,t)$ in the second, and $E_z(x,y,z,t)$ in the third. Similarly it can be shown that B_x, B_y, and B_z each satisfy the classical wave equation, Eq. (104).

The fact that Maxwell's equations yield the classical wave equation (for fields in vacuum) tells us (and told Maxwell) that there must exist electromagnetic waves--standing waves or traveling waves depending on the boundary conditions (i.e., the locations and kinds of antennas, mirrors, etc. that border the vacuum).

Harmonic waves in three dimensions. The three dimensional generalization of the one dimensional harmonic traveling wave

$$\psi(z,t) = A \cos (kz - \omega t) \tag{113}$$

is given by the three-dimensional harmonic traveling plane wave

$$\psi(x,y,z,t) = A \cos (k_x x + k_y y + k_z z - \omega t) \ . \tag{114}$$

Equation (114) satisfies the three-dimensional classical wave equation, Eq. (104), provided we have

$$\omega^2 = c^2(k_x^2 + k_y^2 + k_z^2) \equiv c^2 k^2 \ . \tag{115}$$

The three quantities k_x, k_y and k_z form the three components of a vector, $\underset{\sim}{k}$, called the propagation vector, whose direction is perpendicular to the

wave fronts (the regions of constant phase, which are planes). That is,

$$\text{phase} = k_x x + k_y y + k_z z - \omega t = \text{constant} \qquad (116)$$

is the equation of a plane. To see that Eq. (116) is the equation of a plane and that the vector $\underset{\sim}{k}$ is perpendicular to the plane, we may write the differential of Eq. (116) for constant t .and set it equal to zero. Then

$$k_x\, dx + k_y\, dy + k_z\, dz = 0 \ , \qquad (117)$$

which says that the dot product of $d\underset{\sim}{r}$ and the constant vector $\underset{\sim}{k}$ is zero for any displacement $d\underset{\sim}{r}$ that maintains the condition of constant phase, Eq.(116), at a given time t. The physical meaning of (for example) k_x is obtained by inspection of Eq.(116). If we hold t,y, and z fixed and let x increase by an amount Δx , then the phase increases by an amount $k_x \Delta x$. Therefore k_x is equal to the number of radians of phase per unit distance along x . The other components have analogous meanings.

The phase velocity is given by ω/k , where k is the magnitude of $\underset{\sim}{k}$. That is easily seen by taking the total differential of Eq. (116), allowing t to increase by dt and x, y, and z to increase by dx, dy, and dz in such a way as to maintain constant phase. That gives

$$\omega\, dt - \left(k_x\, dx + k_y\, dy + k_z\, dz\right) = 0 \ , \qquad (118)$$

which says that the dot product of $d\underset{\sim}{r}$ and $\underset{\sim}{k} = k\,\hat{k}$ must equal ωdt to maintain constant phase. Thus $d\underset{\sim}{r} \cdot \hat{k}$, the component of $d\underset{\sim}{r}$ that is perpendicular to the planes of constant phase and is hence in the direction of propagation, must increase by $(\omega/k)dt$ to maintain constant phase. Thus the phase velocity is ω/k . The group velocity of a wave packet also has

the same form as for one-dimensional waves, i.e., v_g is $d\omega/dk$.

The three-dimensional generalization of the one-dimensional standing wave

$$\psi(z,t) = A \sin kz \cos \omega t \tag{119}$$

is (taking a node to be at the origin)

$$\psi(x,y,z,t) = A \sin k_x x \sin k_y y \sin k_z z \cos \omega t \ . \tag{120}$$

The general three dimensional wave can be written as a superposition of traveling waves traveling in all six directions (positive and negative x, y, and z). Alternatively one can write it as a superposition of two standing waves on each axis, with the nodes of one wave on a given axis a quarter wave length away from the nodes of the other wave on that axis. All six standing waves can have arbitrary amplitudes and phase constants. Thus there are six arbitrary amplitudes and six arbitrary phase constants for the general case. Note that one can have a wave that is a traveling wave in some dimensions and a standing wave in others, as (for example)

$$\psi(x,y,z,t) = A \sin k_x x \sin k_y y \cos (\omega t - k_z z) \ , \tag{121}$$

or (for example)

$$\psi(x,y,z,t) = A \sin k_x x \cos (\omega t - k_y y - k_z z) \ . \tag{122}$$

Electromagnetic waves in a dispersive medium. We have already defined the index of refraction for electromagnetic radiation [See Chap. 1, Eq. (82)]

$$n(\omega) \equiv \frac{c}{v(\omega)} \ , \tag{123}$$

and have told you, without proof, that $n(\omega)$ is given by [Chap 1, Eq. (100)]

$$n^2(\omega) = \epsilon(\omega)\mu(\omega), \tag{124}$$

with $\epsilon(\omega)$ and $\mu(\omega)$ defined in esu for a linear, homogeneous, isotropic medium (these terms are defined below) by

$$\epsilon(\omega)\underset{\sim}{E} = \underset{\sim}{E} + 4\pi \underset{\sim}{P} : \quad \epsilon = \text{the dielectric constant} \tag{125}$$

$$\frac{1}{\mu(\omega)} \underset{\sim}{B} = \underset{\sim}{B} - 4\pi \underset{\sim}{M} : \quad \mu = \text{the magnetic permeability} \tag{126}$$

[For Eq. (125) see Vol. 2, p. 323, Eq. (38). To get Eq. (126), see Vol. 2, p. 393, Eq. (55) for the equation $\underset{\sim}{M} = \chi_m \underset{\sim}{H}$, and p. 390, Eq. (52) for the definition $\underset{\sim}{H} = \underset{\sim}{B} - 4\pi \underset{\sim}{M}$. The further definition $\mu = 1 + 4\pi \chi_m$ gives, after a little juggling, Eq. (126).] In this section we will derive the equation for $n(\omega)$, Eq. (124).

We start by writing Maxwell's equations in their most general form (in electrostatic units)

$$\nabla \cdot \underset{\sim}{B} = 0 \tag{127}$$

$$\nabla \cdot \underset{\sim}{E} = 4\pi \rho_{\text{total}} = 4\pi \rho_{\text{free}} - 4\pi \nabla \cdot \underset{\sim}{P} \tag{128}$$

$$\nabla \times \underset{\sim}{B} = \frac{4\pi}{c} \underset{\sim}{J}_{\text{total}} + \frac{1}{c} \frac{\partial \underset{\sim}{E}}{\partial t} = \frac{4\pi \underset{\sim}{J}_{\text{free}}}{c} + 4\pi \nabla \times \underset{\sim}{M} + \frac{4\pi}{c} \frac{\partial \underset{\sim}{P}}{\partial t} + \frac{1}{c} \frac{\partial \underset{\sim}{E}}{\partial t} \tag{129}$$

$$\nabla \times \underset{\sim}{E} = -\frac{1}{c} \frac{\partial \underset{\sim}{B}}{\partial t} \quad . \tag{130}$$

[For Eq. (127) see Vol. 2, page 360; for Eq. (128) see Vol. 2, p. 332, Eq. (57); for Eq. (129) see Vol. 2, p. 343, Eq. (79), which holds when $\underset{\sim}{M}$ is zero, and p. 390, Eq. (50), which holds when $\partial \underset{\sim}{P}/\partial t$ and $\partial \underset{\sim}{E}/\partial t$ are zero; for Eq. (130) see Vol. 2, p. 243, Eq. (30).]

We now consider a linear, homogeneous, isotropic medium. By a linear medium we mean one such that the magnitude of $\underset{\sim}{M}$ is linearly proportional to that of $\underset{\sim}{B}$, and the magnitude of $\underset{\sim}{P}$ is linearly proportional to that of $\underset{\sim}{E}$. (There may be a minus sign in the proportionality factor, i.e., a relative phase difference of 180° for harmonic oscillations; but we do not have anything like, for example, a 90° phase shift between $\underset{\sim}{M}$ and $\underset{\sim}{B}$, or between $\underset{\sim}{P}$ and $\underset{\sim}{E}$. Such a situation can occur, but we are not considering it.) By a homogeneous medium we mean one such that the proportionality factors between $\underset{\sim}{M}$ and $\underset{\sim}{B}$ and between $\underset{\sim}{P}$ and $\underset{\sim}{E}$ are independent of position. Linear relations between $\underset{\sim}{M}$ and

$\underset{\sim}{B}$ and between $\underset{\sim}{P}$ and $\underset{\sim}{E}$ require that $\underset{\sim}{B}$ and $\underset{\sim}{E}$ not be too large. (For sufficiently strong fields, such as can be produced with a pulsed ruby laser, for example, the non-linear contributions to $\underset{\sim}{M}$ and $\underset{\sim}{P}$ can be detected and studied.) By an isotropic medium we mean one such that the vector $\underset{\sim}{M}$ is along the vector $\pm \underset{\sim}{B}$ and the vector $\underset{\sim}{P}$ is along the vector $\pm \underset{\sim}{E}$. (In a crystal, if you push on the electrons of the atoms with force proportional to $\underset{\sim}{E}$, their displacement, which is the source of $\underset{\sim}{P}$, is not necessarily along $\underset{\sim}{E}$, because the "constraint forces" of the crystal make it easier for the electrons to move in some directions than in others.)

The dielectric constant $\epsilon(\omega)$ and magnetic permeability $\mu(\omega)$ depend on frequency and on the details of the charge and current "structure" of the medium. That is what leads to non-trivial dispersion relations--$n(\omega)$ is not independent of ω. Consequently, when we study a particular medium, we will use a harmonic wave with definite frequency ω. The explicit dependence of $\epsilon(\omega)$ and $\mu(\omega)$ on ω will not be considered here. To do that, one must give further detailed specification of the medium, as we did for the case of electromagnetic waves in the ionosphere in Sec. 1.6. In the present section we specify only that the medium is linear, homogeneous, and isotropic. That will make our considerations general, in the sense that they include many different physical situations.

We shall make two additional assumptions, namely, that there be no "free" net charge and no "free" current in the medium. Thus we set:

$$\rho_{free} = 0, \quad \text{and} \quad \underset{\sim}{J}_{free} = 0 \ . \tag{131}$$

Equation (131) does not exclude consideration of a medium like the ionosphere, which has free electrons and free positive ions in equal number. The motion of the charges in a neutral medium will be included under "polarization."

Finally we are ready to proceed with the derivation of the index of refraction, Eq. (124). We stop thinking and do algebra. We start with Maxwell's equations [augmented by Eq. (131)]. The first equation, $\underset{\sim}{\nabla} \cdot \underset{\sim}{B} = 0$, is unaffected by any known medium. (This means that no magnetic monopoles have yet been discovered.) After insertion of Eq. (131), the second equation, Eq. (128), can be written

$$\underset{\sim}{\nabla} \cdot \left\{ [\underset{\sim}{E} + 4\pi \underset{\sim}{P}] \right\} = 0 \quad , \text{ i.e., } \underset{\sim}{\nabla} \cdot (\epsilon \underset{\sim}{E}) = 0 \tag{132}$$

Similarly, Eq. (129) can be written in the form

$$\underset{\sim}{\nabla} \times \left\{ \underset{\sim}{B} - 4\pi \underset{\sim}{M} \right\} = \frac{1}{c} \frac{\partial}{\partial t} \left\{ \underset{\sim}{E} + 4\pi \underset{\sim}{P} \right\} ,$$

i.e.,

$$\underset{\sim}{\nabla} \times (\underset{\sim}{B}/\mu) = \frac{1}{c} \frac{\partial}{\partial t} (\epsilon \underset{\sim}{E}) \quad . \tag{133}$$

Equation (127) is unaffected by the presence of the medium but for convenience it can be written in the form

$$\underset{\sim}{\nabla} \cdot (\underset{\sim}{B}/\mu) = 0 \quad , \tag{134}$$

because the medium is homogeneous, i.e., μ is constant in space. Likewise Eq. (130) is unchanged by the medium, but for convenience [i.e., to have $\epsilon \underset{\sim}{E}$ and $\mu^{-1} \underset{\sim}{B}$ appear, instead of $\underset{\sim}{E}$ and $\underset{\sim}{B}$] we make use of the homogeneity to write it in the form

$$\underset{\sim}{\nabla} \times (\epsilon \underset{\sim}{E}) = - \frac{\mu\epsilon}{c} \frac{\partial (\underset{\sim}{B}/\mu)}{\partial t} \quad . \tag{135}$$

For ease of reference we now collect our results [Eqs. (132), (133), (134), and (135)] all in one place:

$$\nabla \cdot (\underset{\sim}{B}/\mu) = 0 \tag{136a}$$

$$\nabla \cdot (\epsilon \underset{\sim}{E}) = 0 \tag{136b}$$

$$\nabla \times (\underset{\sim}{B}/\mu) = \frac{1}{c} \frac{\partial(\epsilon \underset{\sim}{E})}{\partial t} \tag{136c}$$

$$\nabla \times (\epsilon \underset{\sim}{E}) = -\frac{\mu\epsilon}{c} \frac{\partial(\underset{\sim}{B}/\mu)}{\partial t} \tag{136d}$$

Equations (136) are Maxwell's equations for a linear, homogeneous, isotropic, neutral, current free medium. [Note that, since μ and ϵ are assumed constant, there are other equivalent ways to write Eqs. (136).]

Now that we have the appropriate form of Maxwell's equations, we proceed in the same way that we did in the preceding example, where we derived the classical wave equation for electromagnetic waves in vacuum. This time we shall not find the classical wave equation, but will find the equation

$$\frac{\partial^2 \psi(x,y,z,t)}{\partial t^2} = \frac{c^2}{\mu(\omega)\epsilon(\omega)} \nabla^2 \psi(x,y,z,t) \quad , \tag{137}$$

where $\psi(x,y,z,t)$ can be any one of the six wave functions that are the components of $\epsilon \underset{\sim}{E}$ and $\underset{\sim}{B}/\mu$ (or alternatively the components of $\underset{\sim}{E}$ and $\underset{\sim}{B}$). Equation (137) is not the classical wave equation unless $\mu(\omega)\epsilon(\omega)$ is independent of ω. [See the discussion of Eqs. (86) and (87).]

To obtain Eq. (137) for $\epsilon \underset{\sim}{E}$ we differentiate Eq. (136c) once with respect to time, interchange the order of the time and space derivatives, and then use Eq. (136d). Thus

$$\frac{\partial^2(\epsilon\underset{\sim}{E})}{\partial t^2} = c\frac{\partial}{\partial t}[\underset{\sim}{\nabla}\times(\underset{\sim}{B}/\mu)] = c\underset{\sim}{\nabla}\times\left[\frac{\partial}{\partial t}(\underset{\sim}{B}/\mu)\right] = -\frac{c^2}{\mu\epsilon}\underset{\sim}{\nabla}\times[\underset{\sim}{\nabla}\times(\epsilon\underset{\sim}{E})]. \quad (138)$$

Using the identity for curl curl $\underset{\sim}{C}$ given by Eq. (111), with $\underset{\sim}{C}$ taken to be $\epsilon\underset{\sim}{E}$, using $\underset{\sim}{\nabla}\cdot(\epsilon\underset{\sim}{E}) = 0$ [Eq. (126b)], and writing the symbol ∇^2 for $\underset{\sim}{\nabla}\cdot\underset{\sim}{\nabla}$, we obtain

$$\frac{\partial^2(\epsilon\underset{\sim}{E})}{\partial t^2} = \frac{c^2}{\mu\epsilon}\nabla^2(\epsilon\underset{\sim}{E}) \quad . \quad (139)$$

This vector equation means the same as equations of the form of Eq. (137) for each of the three components ϵE_x, ϵE_y, and ϵE_z. Similarly, we can start with Eq. (136d), differentiate with respect to time, and use Eq. (136c) to obtain (letting you do the arithmetic this time)

$$\frac{\partial^2(\underset{\sim}{B}/\mu)}{\partial t^2} = \frac{c^2}{\mu\epsilon}\nabla^2(\underset{\sim}{B}/\mu) \quad . \quad (140)$$

We see that Eqs. (139) and (140) are of the form of Eq. (137), which is the three-dimensional generalization of [see Eq. (86)]

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = v^2(\omega)\frac{\partial^2\psi(z,t)}{\partial z^2} \quad . \quad (141)$$

with

$$v^2(\omega) \equiv \frac{c^2}{n^2(\omega)} = \frac{c^2}{\mu(\omega)\epsilon(\omega)} \quad . \quad (142)$$

Thus

$$n^2(\omega) = \mu(\omega)\,\epsilon(\omega) \quad , \quad (143)$$

which is Eq. (124).

When we write Eq. (141), we are considering a single frequency ω, so we might as well replace $\partial^2\psi/\partial t^2$ by $-\omega^2\psi$. This gives

$$\frac{\partial^2\psi(z,t)}{\partial z^2} = -\frac{\omega^2\mu(\omega)\epsilon(\omega)}{c^2}\psi(z,t) \quad . \tag{144}$$

When $\omega^2\mu(\omega)\epsilon(\omega)/c^2$ is positive, we call it k^2 :

$$k^2 = \frac{\omega^2\mu(\omega)\epsilon(\omega)}{c^2} \quad , \quad k^2 > 0 \quad . \tag{145}$$

The general solution of Eq. (141) or (144) for positive k^2 can be written in any one of the following equivalent forms (which are merely different linear combinations of one another):

$$\psi(z,t) = A \cos kz \cos \omega t + B \cos kz \sin \omega t + C \sin kz \cos \omega t + D \sin kz \sin \omega t \tag{146}$$

$$= A' \cos kz \cos (\omega t + \phi_1) + B' \sin kz \cos (\omega t + \phi_2) \tag{147}$$

$$= A'' \cos (\omega t - kz + \phi_1') + B'' \cos (\omega t + kz + \phi_2') \quad . \tag{148}$$

Notice that Eq. (147) is a superposition of two standing waves that differ by $\frac{1}{4}$ cycle in space and have arbitrary phase constants. Equation (148) consists of a superposition of two traveling waves, traveling in opposite directions with arbitrary phase constants. Equation (146) is non-committal. It includes all possibilities without "prejudice" as to whether one has standing or traveling waves or some combination of both kinds.

When $\omega^2\mu(\omega)\epsilon(\omega)/c^2$ is negative, we call it $-|k|^2$:

$$-|k|^2 = \frac{\omega^2\mu(\omega)\epsilon(\omega)}{c^2} \quad , \quad |k|^2 > 0 \quad . \tag{149}$$

(One may instead simply keep calling $\omega^2\mu(\omega)\epsilon(\omega)/c^2$ by the name k^2 when it is negative, and then say that k is a "pure imaginary" quantity $+i\,|k|$ or $-i\,|k|$. We prefer to avoid mentioning $\sqrt{-1}$, so as to avoid

having "physical" quantities that involve $\sqrt{-1}$. No one ever found a wave with wavelength $\sqrt{-1}$ centimeters. But an exponential damping distance of 1 centimeter is perfectly ordinary.) In this case the general solution of the wave equation, Eq. (141), can be written in either of the following equivalent forms:

$$\psi(z,t) = A \cosh |k|z \cos \omega t + B \cosh |k|z \sin \omega t + C \sinh |k|z \cos \omega t + D \sinh |k|z \sin \omega t \qquad (150)$$

$$= A' e^{+|k|z} \cos (\omega t + \phi_1) + B' e^{-|k|z} \cos (\omega t + \phi_2) \quad . \qquad (151)$$

[The functions cosh and sinh are defined in the Appendix, Eqs. (7) through (12).] These solutions correspond to oscillations of frequency ω with magnitude exponentially changing with distance. This is the situation we found in Sec. 1.6 for waves incident on the ionosphere with frequency lower than the plasma frequency.

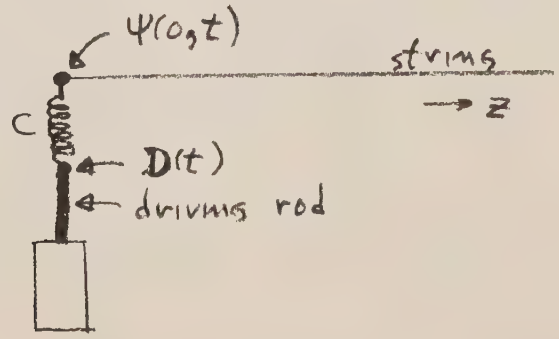
Esoteric examples. In Advanced Topic 4 we obtain the Schrödinger equation for non-relativistic de Broglie waves and the Klein-Gordon equation for relativistic free particles.

In Advanced Topic 5 we find the three-dimensional form of the classical wave equation, Eq. (104), for the case of spherical symmetry, and show that the outgoing-spherical-wave function $\psi(r,t) = A r^{-1} \cos (\omega t - kr)$ Eq. (8) is a solution.

Problems, Chapter 2.

1. Modify the derivation of the dispersion relation for the beaded string given in Sec. 1.6 so that you end up with traveling harmonic waves rather then standing harmonic waves. (See the discussion starting two sentences before Eq. (4), Chap. 2.)

*2. Suppose that instead of tying the semi-infinite string directly to the output of the driving mechanism (as in Fig. 2-1) it is coupled to the transmitter through a spring as follows:



The string tension is T, the mass density of the string is ρ, and the spring constant is C. The length of the spring is such that if $D(t)$ is zero and the spring is relaxed then $\psi(0,t)$ is zero. The motion of the rod is given by $D(t) = A \cos \omega t$. Assume there is a traveling wave of the form $\psi(z,t) = B \cos (\omega t - kz + \phi)$. The problem is to figure out what the "boundary condition" is at $z = 0$, and then apply it to find B/A and ϕ. Hint: The algebra is easier if you use complex numbers.

Answer: $\tan \phi = -\omega(T\rho)^{\frac{1}{2}}/C$, $B/A = [1 + (\omega^2 T\rho/C)[1 + (\omega^2 T\rho/C^2)]^{-\frac{1}{2}}$ $= \cos \phi$. Notice that for C huge we get $\psi(0,t) = D(t)$, as we should expect. Why?

3. Show that the sum of two traveling harmonic waves $A_1 \cos(\omega t - kz + \phi_1)$ and $A_2 \cos(\omega t - kz + \phi_2)$ (i.e., both traveling in the +z direction, and having the same frequency ω) is itself a harmonic traveling wave of the same kind. That is, the sum can be written in the form $A \cos(\omega t - kz + \phi)$. Find how A and ϕ are related to A_1, A_2, ϕ_1, and ϕ_2. Hint: The use of complex numbers or a rotating vector diagram helps immensely.

4. Plug numbers into Eqs. (50) and (51) for water, using $T = 72$ and $g = 980$ (cgs units), to find (a)the wave length for which the phase and group velocities are equal; (b)the phase and group velocities for ocean waves of wavelength 20 feet; 200 feet; (c) the phase and group velocities for waves of 2 mm wavelength. When you throw a stone into a pond, you will see (if you look carefully) small amplitude, millimeter (or so) wavelength "surface tension waves" spread rapidly away from the disturbance. The longer wavelength (a few cm) circular waves, which make up the major portion of the wave group, travel slower than the surface tension components.

5. Consider a wave group of gravitational deep-water waves (with $\omega^2 = gk$) such that at $t = 0$ the length of the group is $(\Delta z)_o$. The dominant wavelength of the group is λ_o. Show that after the group has traveled a distance Z it is longer than $(\Delta z)_o$ by an amount such that the group contains $Z/(\Delta z)_o$ more half waves (of the dominant wavelength) than it did at $t = 0$, i.e., the added length is roughly $[Z/(\Delta z)_o](\lambda_o/2)$. Hint: See Eq. (78).

6. Consider electromagnetic radiation in a linear, homogeneous, isotropic medium. Suppose the magnetic permeability μ is 1. Then $n(\omega) = [\epsilon(\omega)]^{\frac{1}{2}}$. according to the theory of relativity, no signal can propagate faster than

$c = 3.0 \times 10^{10}$ cm/sec. What limitation does that put on the possible variation of $\epsilon(\omega)$ with ω ? (Assume $\epsilon(\omega)$ is positive for all ω .)

Answer: $\omega \frac{dn}{d\omega} + (n-1) \geq 0$.

7. Show that in any non-dispersive medium the phase and group velocities are equal.

8. Green light has wavelength in vacuum about 5500 $\mathring{A}$. What is the frequency in Hz? (One Hz = 1 Hertz = 1 cps.) What is the wavelength in glass having index of refraction $n = 1.5$? What is the frequency in the glass? What is the wave number $k/2\pi$ in cycles/cm (usually called inverse centimeters) in vacuum? In the glass?

9. Suppose that a point a on a string at $z_a = 10$ cm oscillates in harmonic motion at frequency 10 cps with amplitude 1 cm. Its phase is such that at $t = 0$ the point on the string is passing through its equilibrium position with upward velocity (positive displacement is upward). (a) What is the magnitude and direction of the velocity of point a at $t = 0.05$ sec? Suppose the string parameters (mass per unit length and tension) are such that the wave velocity is 100 cm/sec. (b) What is the wavelength of a traveling wave? What is the wavelength of a standing wave? (c) Another point b at $z_b = 15$ cm oscillates with the same amplitude as that at $z_a = 10$ cm, but with a relative phase of 180° with respect to the oscillation at z_a . Can you tell whether we have here a pure traveling wave, a pure standing wave, or a combination? (d) A third point c at 12.5 cm also oscillates with the same amplitude as that at z_a but 180° out of phase with point a . Now tell us whether the wave is a traveling or a standing wave (or a combination).

10. Home experiment. Measure the approximate bandwidth broadcast and received in your AM radio by twisting the tuning dial and seeing what the extremes are for receiving a given station. (The highest reading on the AM dial is usually 130. This means 1300 Kc.) How does your result compare with the result which we obtained in Sec. 2.4, namely that one needs $\Delta\nu \approx 40$ Kc to cover both side bands for high fidelity sound reproduction?

11. Tubas can play very low, for example C_1 32.7 cps (the lowest C on the piano, called C_1). Flutes can play very high, their highest note normally being C_7 (2093) (one octave below the highest note on the piano). Each note on the equal-tempered scale differs from its neighbor by a factor of about 1.06. Flutes can play very fast. Tubas don't go so fast. Is that the fault of the tuba player? Or the tuba? Could the tuba be re-designed so that tuba players could play as fast as flutists? What would you calculate as reasonable top speed for tubists playing scales near C 32.7? For flutists playing near C 2093? First you have to decide on a reasonable musical criterion, then do physics.

Answer: Two notes per second for tuba; 120 per second for flute (wow!).

12. A man brought his AM radio into a repair shop and complained that the tuning wasn't fine enough. He wanted a given station to be very sharply defined on the dial. So they fixed it according to his desires. Then he brought it back again. What was his complaint this time?

13. Fourier analyzing machine. Suppose we wish to know what frequencies are present in a force F(t) which acts from time t = 0 to time t = t_1. (The force may be acting all the time, but any experiment done by people has to start sometime and end sometime. The time t = 0 is when we turn our machine on. The time t_1 is the time when we decide to look at the data and see what the machine has to tell us. We can let t_1 vary but will keep t = 0 fixed as the turn on time of the machine.) Our machine consists of a very large number of frictionless pendulums (or linear harmonic oscillators, so that we do not have to worry about the non-linear behavior of a pendulum for large amplitudes; we shall still call them pendulums). The angular frequency for free oscillations of a given pendulum will be called ω. Each pendulum has a different natural frequency ω and we have as many pendulums as are necessary to cover uniformly whatever frequency interval is of interest. We somehow design a method to apply the force F(t) to each pendulum. At time t = 0 we arrange to have all of the pendulums hanging at their equilibrium positions, with zero velocity. At time t_1 we look at the oscillation amplitude and phase of each pendulum. It is this collection of information, the amplitudes and phases of the pendulums, which is the entire output of the machine. The machine might perhaps be designed so that it would display for us a plot of the square of the amplitude of the pendulum having frequency ω (or of all the pendulums having frequencies in some small interval $\Delta\omega$ around the frequency ω, if we have a huge number of pendulums distributed almost continuously in ω) as a function of ω, and also a plot of the phase constant as a function of ω. Notice that if the force F(t) has died away before the time t_1 then during the time between the dying of F(t) and the read-out time t_1 the amplitudes of all the pendulums remains constant (they are frictionless), but their relative phases continue to change because they have different natural frequencies. All of this can be taken into account in analyzing the output at time t_1. Furthermore,

there may be the situation where we want to reset the pendulums to zero as soon as we have arrived at time t_1 and let them run for another time interval from t_1 to $2t_1$, then $2t_1$ to $3t_1$, etc. That will be true in experimental situations where the force $F(t)$ that we wish to study is "triggered" by some means under our control and is always essentially "dead" by a time t_1 after the trigger pulse. We can then repeat the triggering and analysis over and over.

Let us leave these engineering details and turn to the physics (the easy problems). Consider one pendulum of natural frequency ω. Newton's second law gives

$$\frac{md^2x(t)}{dt^2} = -m\omega^2 x(t) + F(t) , \qquad (P1)$$

which is to be solved with the conditions that $x(t)$ and dx/dt are zero at time zero. We shall not tell you how to solve for the general case. (That is discussed in Advanced Topic 7.) Instead we choose a particular example. Suppose that $F(t)$ is given by

$$F(t) = F_o \cos \omega_o t, \quad \text{for } 0 \leq t \leq t_1,$$

$$F(t) = 0 \quad \text{for } t < 0 \text{ and for } t > t_1 . \qquad (P2)$$

We have given the name t_1 to the time when the force turns off, because we might as well read the pendulums at that time. (If we wish to read them later we can always extrapolate their phases back to time t_1.) Thus we have

$$\frac{d^2x(t)}{dt^2} = -\omega^2 x(t) + m^{-1} F_o \cos \omega_o t \qquad (P3)$$

in the indicated time interval with the conditions that x and dx/dt are zero at $t = 0$. How can we solve this differential equation? Let us guess. There are two frequencies involved: one is ω_o, the other is ω. Let us try a

superposition of harmonic oscillations A cos $\omega_o t$, B sin $\omega_o t$, C cos ωt and D sin ωt with unknown constants A,B,C,D to be determined. If we can find a solution that satisfies all of the conditions of the problem, then we must have the unique solution. (Either your physical intuition or a study of this kind of differential equation should tell you that there is only one solution--the pendulum always behaves in a prescribed manner; we have left it no freedom of choice.)

Finally it is your turn to work:

(a) Show that C must equal -A.

(b) Show that ωD must equal $-\omega_o B$.

(c) By using Eq. (P3) show that for our particular choice of phase constants, namely that F(t) is proportional to cos $\omega_o t$ (and not, for example, sin $\omega_o t$), we must have B = 0.

(d) Show that the solution is

$$x(t_1) = m^{-1}F_o \frac{[\cos \omega t_1 - \cos \omega_o t_1]}{[\omega_o^2 - \omega^2]} \qquad (P4)$$

(e) Show that $x(t_1)$ can be written in the form

$$x(t_1) = A(\omega,\omega_o,t_1) \sin \omega_{av} t_1 \, , \qquad (P5)$$

$$\omega_{av} = \frac{1}{2}(\omega_o + \omega) \, , \qquad (P6)$$

$$A(\omega,\omega_o,t_1) = \frac{m^{-1}F_o}{2\omega_{av}} \frac{\sin \frac{1}{2}(\omega - \omega_o)t_1}{\frac{1}{2}(\omega - \omega_o)} \, . \qquad (P7)$$

(f) Show that dx/dt at $t = t_1$ can be obtained by differentiating Eq. (P4) with respect to t_1. Since we know x and dx/dt at time t_1 for all pendulums (all ω), we can let the pendulums swing freely after t_1 (when F becomes zero) until we are ready to read them and then extrapolate our readings back to time t_1, if we wish.

A plot of the amplitude $A(\omega,\omega_o,t_1)$ versus ω gives us a large part of the information about the Fourier spectrum. [The entire information is contained in Eqs. (P5), (P6) and (P7).] We notice the following features(show them to be true):

1. $A(\omega,\omega_o,t_1)$ is largest in magnitude at $\omega = \omega_o$, where it has the value

$$A(\omega_o,\omega_o,t_1) = \frac{m^{-1}F_o}{2\omega_o} \quad t_1 \tag{P8}$$

2. $A(\omega,\omega_o,t_1)$ goes to zero at frequencies ν_+ and ν_-, where

$$(\nu_+ - \nu_o)t_1 = 1, \quad (\nu_- - \nu_o)t_1 = -1 . \tag{P9}$$

3. A plot of the square of $A(\omega,\omega_o,t_1)$ versus ω shows that A^2 is down to about half of its maximum value when ν is given by $\nu = \nu_o + \frac{1}{2} t_1^{-1}$ or by $\nu = \nu_o - \frac{1}{2} t_1^{-1}$. Therefore if we define the "full width at half maximum" of the plot of A^2 versus ν to be the "width" (in units of frequency) between the two "points" (on the plot) where A^2 is half of its maximum value, and if we give this full width at half maximum the name $\Delta\nu$, and give t_1 the name Δt, show that

$$\Delta\nu \; \Delta t = 1. \tag{P10}$$

4. Show that the "area" under a plot of $A(\omega,\omega_o,t_1)$ versus ω remains constant as the time t_1 increases. (By "area" we mean the product of A times $d\nu$ integrated over all pendulum frequencies ν.) Show that this is plausible by considering the region of the plot near $\nu = \nu_o$. If you are good at integration show it exactly. (This is optional and not recommended. We advise, for a first try, neglecting the dependence of ω_{av} on ω in doing the integral. Then you can find the appropriate integral in a table.) Thus the plot becomes more and more sharply concentrated at $\nu = \nu_o$ as t_1 increases. In the limit that t_1

becomes huge, $A(\omega,\omega_o,t_1)$ is proportional to what is called a "delta function" of the frequency difference $\omega - \omega_o$. In order to convert A into a "delta function of $\omega - \omega_o$," we need only renormalize it so that its "area" on a plot of A versus ω is unity and take the limit where t_1 is bigger than any physically significant time in the problem. We call such a time by the nickname "infinity," but, since this is physics, we never mean whatever it is that mathematicians mean by that word.

5. Show that the area under a plot of $A^2(\omega,\omega_o,t)$ versus ω increases linearly with time t_1. That shows that the total energy stored in <u>all</u> of the pendulums increases linearly with the time t_1. As time goes on this energy becomes (fractionally) more and more concentrated into the pendulums with ν near ν_o. Show that the energy of the pendulum with frequency exactly at ν_o increases as the square of the time t_1, but that the total number of "important" pendulums (near ν_o) goes as t_1^{-1}, so that the total energy in the pendulums goes as t_1.

In this problem we have shown you how to design a Fourier analyzing machine (although we left the engineering details to you) and have shown you what you would find for the case of a force F(t) given by Eqs. (P2). The result illustrates the general principle that for a pulse with time duration approximately Δt and with dominant frequency ν_o, the Fourier spectrum consists of a spread of important frequencies lying in a band of width $\Delta\nu$, centered at $\nu = \nu_o$, and with band width $\Delta\nu \approx (\Delta t)^{-1}$. The exact details and the amounts and phases of frequency contributions <u>outside</u> of the dominant band Δt depend on the specific function F(t). In this example we found the exact solution for the Fourier spectrum, and of course it includes all conceivable details as well as satisfying the general results $\Delta\nu\ \Delta t \approx 1$.

Chapter 3 - Emission and absorption of waves: Impedance.

Section		Page
3.1	Introduction	1
3.2	Emission and absorption of traveling waves on a continuous string.	8
3.3	Sound waves.	22
*3.4	Current and voltage waves on a transmission line.	47
*3.5	Radiation by an oscillating sheet of charge.	65
3.6	The general case of electromagnetic plane waves in vacuum.	72
*3.7	Reflection at a reactive load.	80
	Problems and Home Experiments.	98

3.1 Introduction.

In this chapter we shall examine the "mechanism" by which traveling waves are emitted by a transmitter and that by which they are reflected when they encounter a discontinuity in the medium. We shall study four examples -- transverse waves on a continuous string (Sec. 3.2), longitudinal sound waves in a tube (Sec. 3.3), transverse electromagnetic waves in a parallel plate transmission line (Sec. 3.4), and transverse electromagnetic waves emitted by an oscillating sheet of charge (Sec. 3.5). [Sections 3.4 and 3.5 are optional examples. They can be omitted without loss of continuity.]

In every case the transmitter has some moving "output" part which moves with a velocity which we shall call $u(t)$:

$$u(t) = \text{velocity of transmitter output} \quad . \tag{1}$$

In the case of the string, $u(t)$ is the transverse velocity of the oscillating output connection of the transmitter. In the case of sound waves, $u(t)$ is the velocity of the piston which oscillates in longitudinal motion at one end of the tube. In the case of electromagnetic waves in a transmission line, $u(t)$ is the velocity of the oscillating charges at the output terminals of the source of voltage which drives the line. This velocity is proportional to the electric current $I(t)$ at the output terminals, and we shall use the current $I(t)$ instead of velocity $u(t)$ as the quantity of interest. In the case of the oscillating sheet of charge (which may be thought of as being in a large flat antenna), the "output" moving part (i.e., the moving part that is coupled to the waves) is the charge. Again we take current $I(t)$ instead of velocity $u(t)$ to be the interesting quantity.

In every case we shall find that when traveling waves are being emitted into an open medium the driving mechanism of the transmitter -- whatever it

is that drives the output part at velocity $u(t)$ [or current $I(t)$] -- has to push on the output part with a force that is linearly proportional to the velocity $u(t)$ [or current $I(t)$]. In the electromagnetic examples the driving force is provided by an electric field produced by the driving mechanism which pushes the charges. Now, an electromotive force V is proportional to an electric field (it is given by electric field times a distance), and in the electromagnetic examples we use electromotive force $V(t)$ to represent the force driving the charges at the transmitter output. The proportionality constant relating the driving emf V and the resulting current I is called the characteristic impedance (or, for short, just impedance). In the two electromagnetic examples, the characteristic impedance, which we shall denote by Z, has dimensions of resistance. Thus

$$V(t) = Z\, I(t) \quad . \tag{2}$$

In the mechanical examples we shall use the same name, characteristic impedance, and the same symbol, Z, to represent the proportionality constant between the driving force $F(t)$ and the output velocity $u(t)$. Thus we have

$$F(t) = Z\, u(t) \quad . \tag{3}$$

Of course the dimensions of Z differ between the mechanical and electromagnetic cases, but that will not cause us any difficulty.

The characteristic impedance Z depends on the physical processes involved in the wave emission, i.e., on the coupling of the transmitter to the medium, and on the properties of the medium. In all the processes we shall consider, we shall find that the characteristic impedance depends on the same two properties of the medium as does the propagation velocity v,

namely the "inertia-like" property and the "return-force-like" property. In fact, the propagation velocity v and the impedance Z may be regarded merely as a way to re-express the inertia and return force. In the case of electromagnetic waves, the velocity and impedance have the desirable property that they are directly measurable quantities, whereas the "inertia" and the "return force" are only recognizable by analogy with mechanical examples.

It is important to realize that a characteristic impedance Z satisfying Eq. (2) or Eq. (3) exists only because the transmitter is coupled to an open medium and is emitting pure outgoing traveling waves. The open medium then acts like a "resistive load impedance" (the word resistive is used because Eq. (2) is like Ohm's law, with Z the resistance. "Resistive" is to be contrasted with "capacitive" or "inductive," i.e., "reactive"). In fact, the driving mechanism cannot tell whether it is actually radiating waves into an open medium or merely driving a "lumped" resistive load. Thus if you were to disconnect the oscillator output from the antenna of your local radio station and connect the oscillator across an equivalent resistor (one having the same resistance Z as that which the oscillator experiences when it is driving the electrons in the antenna), then the oscillator could not tell the difference. In the case of the continuous string, if you were to replace the string (which is being shaken by the oscillating rod of the transmitter) by a suitable "dashpot," the transmitter would experience the same drag force as when it is emitting traveling waves into an infinite string.

If the medium is not open, so that reflected waves return and add their contributions to the force on the transmitter output part, the net force

exerted by the waves on the transmitter output part is not a pure resistive drag. For a resistive drag the forces exerted by the waves on the transmitter have a phase of 180° with respect to the transmitter output velocity $u(t)$. If reflected waves are present, the forces they exert on the transmitter will not, in general, have this phase relation. The load on the transmitter is then called "reactive." Force components are called "capacitive" or "inductive" depending on whether they lead the output velocity $u(t)$ by 90° or lag by 90°, respectively.

The radiated power, denoted by $P(t)$, is given for the mechanical examples by the product of the driving force $F(t)$ and the velocity $u(t)$ of the output part of the transmitter; for the electromagnetic examples, by the product of the driving electromotive force $V(t)$ and the current $I(t)$ at the transmitter output:

$$\text{mechanical:} \quad P(t) = F(t)u(t) \tag{4}$$

$$\text{electromagnetic:} \quad P(t) = V(t)I(t) \tag{5}$$

These relations are general, not depending on purely resistive load. In case the medium is open (i.e., no reflected waves return to the transmitter), the medium provides a resistive drag given by Eq. (3) or Eq. (2). Then we have

$$\text{mechanical:} \quad P(t) = F(t)u(t) = Zu^2(t) = F^2(t)/Z \tag{6}$$

$$\text{electromagnetic:} \quad P(t) = V(t)I(t) = ZI^2(t) = V^2(t)/Z \,, \tag{7}$$

[If a pure standing wave is coupled to a transmitter at $z = 0$, the force exerted on the transmitter output part by the wave is purely reactive. Then the average transmitter output power, $\langle P(T) \rangle$, is zero. (In this notation for averages the argument T denotes that the average is taken over one period T.) However, during the cycle some energy may flow in and out of the transmitter, so that $P(t)$ is not zero at every instant but is sometimes positive

and sometimes negative.]

In each of our examples we shall make use of Newton's laws (and Maxwell's equations, where necessary) to find the relation between the transmitter output velocity u(t) and the radiated waves (i.e., the emitted traveling waves). Then we shall be able to use Eqs. (6) or (7) to express the energy flux in a traveling wave in terms of the wave amplitudes and their derivatives with respect to space and time.

We shall also learn that we can achieve "perfect termination" of a traveling wave by terminating the medium in a lumped resistive load that has the same impedance as the characteristic impedance Z of the medium. It is clear that if the transmitter cannot tell whether it is connected to an open wave-carrying medium or to a lumped resistive load, neither can the medium itself at some point z tell whether it is connected to more of the medium (extending to infinity) or to a termination with impedance Z. In either case the traveling wave will be absorbed without reflection. In the case where there is an infinite continuation of the medium beyond z the wave is not really absorbed, but the point at z cannot tell the difference. See Fig. 1.

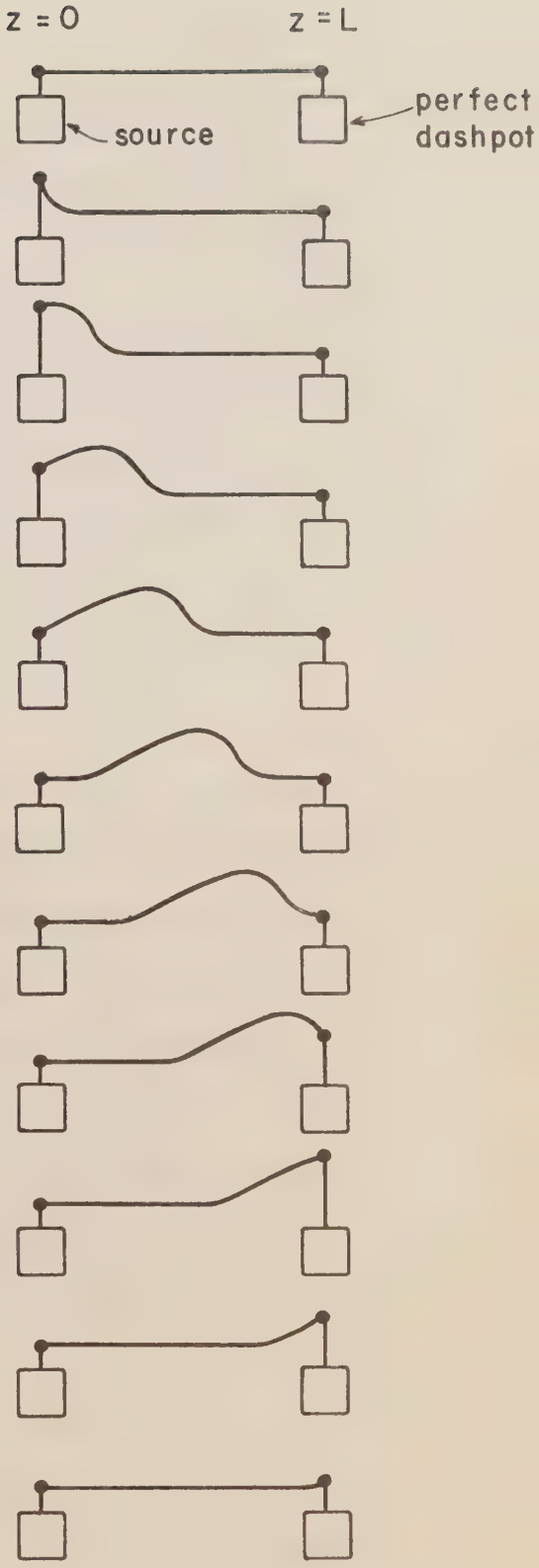


Fig. 3-1. The source cannot tell the length of the string if the string is perfectly terminated. As far as the source can tell, it may be connected to an infinitely long string, or it may be connected directly to the input terminal of the dashpot. Actually it is here connected to a finite string that is perfectly terminated.

When the medium is terminated with a resistive load impedance Z_L that does not match the characteristic impedance, a reflected wave is generated. Again, it is impossible for the wave to tell a "lumped" resistive load Z_L from an extended wave-carrying medium of characteristic impedance Z_L. We shall find that the amplitude of the reflected wave is proportional to that of the incident wave and to the "fractional mismatch" of impedances. Thus the ratio of reflected to incident amplitude, called the reflection coefficient R, is given by

$$R_{12} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad , \tag{8}$$

where the subscript 12 means "from 1 to 2," i.e., incident from medium 1 of impedance Z_1 onto load impedance Z_2, where Z_2 may be either a lumped impedance or the characteristic impedance of medium 2.

We shall introduce all of these concepts with the aid of the stretched string in Sec. 3.2. Then we shall go on to the more interesting examples of sound waves and two cases of electromagnetic waves in Secs. 3.3, 3.4, and 3.5. In Sec. 3.6 the results of Secs. 3.4 and 3.5 for the two electromagnetic cases considered will be shown to hold generally for electromagnetic waves.

In Sec. 3.7 we shall consider what happens when a traveling wave is incident on a reactive load impedance or on an extended reactive medium (i.e., a medium which offers a reactive impedance to the incident wave at the boundary. We encountered this situation in treating a plasma with incident light of $\omega < \omega_p$). Then we shall find that the wave function in the reactive medium has the form

$$\psi(z,t) = A\, e^{-|k|\, z} \cos \omega t \quad , \tag{9}$$

where the reactive medium extends from $z = 0$ to $+\infty$. (See Sec. 2.5.) In the dispersive medium (from $z = -\infty$ to $z = 0$) the wave is a superposition of an incident traveling wave given by

$$\psi_{inc}(z,t) = A \cos(\omega t - kz) \tag{10}$$

and a reflected wave given by

$$\psi_{ref}(z,t) = A \cos(\omega t + kz + \alpha) \tag{11}$$

where the phase shift constant α may be thought of as crudely corresponding to a time delay during which the incident wave enters the reactive medium, penetrates for a distance of order $|k|^{-1}$, turns around and comes back as the reflected wave. The reflected wave has the same amplitude A as the incident wave -- no energy is permanently lost to the reactive medium, if it is thick enough.

If the reactive medium is not infinitely thick, one has a phenomenon called barrier penetration. In that case the wave function in the reactive medium contains a contribution like that of Eq. (9) and another similar contribution with $e^{+|k|z}$ instead of $e^{-|k|z}$. If the reactive medium ends at $z = L$, and if a new dispersive medium (i.e., one that will support traveling waves of the incident frequency) extends from L to $+\infty$, then in the new region $\psi(z,t)$ will have the form of a traveling wave propagating in the $+z$ direction (assuming the only source of waves is at $z = -\infty$). Thus for this case, $\psi(z,t)$ will have the form (valid from $z = L$ to $+\infty$)

$$\psi(z,t) = B \cos(\omega t - kz + \beta) ,$$

where the amplitude B depends on the thickness and properties of the reactive medium. (Note that B must satisfy the relation $|B| < |A|$.) The phase constant β corresponds to a time delay to get through the reactive medium, including the effect of multiple traversals due to partial reflections at each boundary.

Section 3.7 can be omitted without loss of continuity.

3.2 Emission and absorption of traveling waves on a continuous string.

First we consider the details of the emission process. We assume the string is perfectly continuous and has tension T and mass per unit length ρ. We also assume that the slope is always small so that we have linear relations and can use small-angle approximations. We do not assume that we know the relation between the wave propagation velocity v and the string parameters T and ρ.

For simplicity let us consider the case that $F(t)$, the force exerted by the transmitter on the string at $z = 0$, is a step function. The force is turned on at time zero and is constant after that. The string is attached to the transmitter in such a way that $\psi(0,t) = D(t)$, the displacement of the transmitter output hook. The wave traveling down the string then has the simple form shown in Fig. 2.

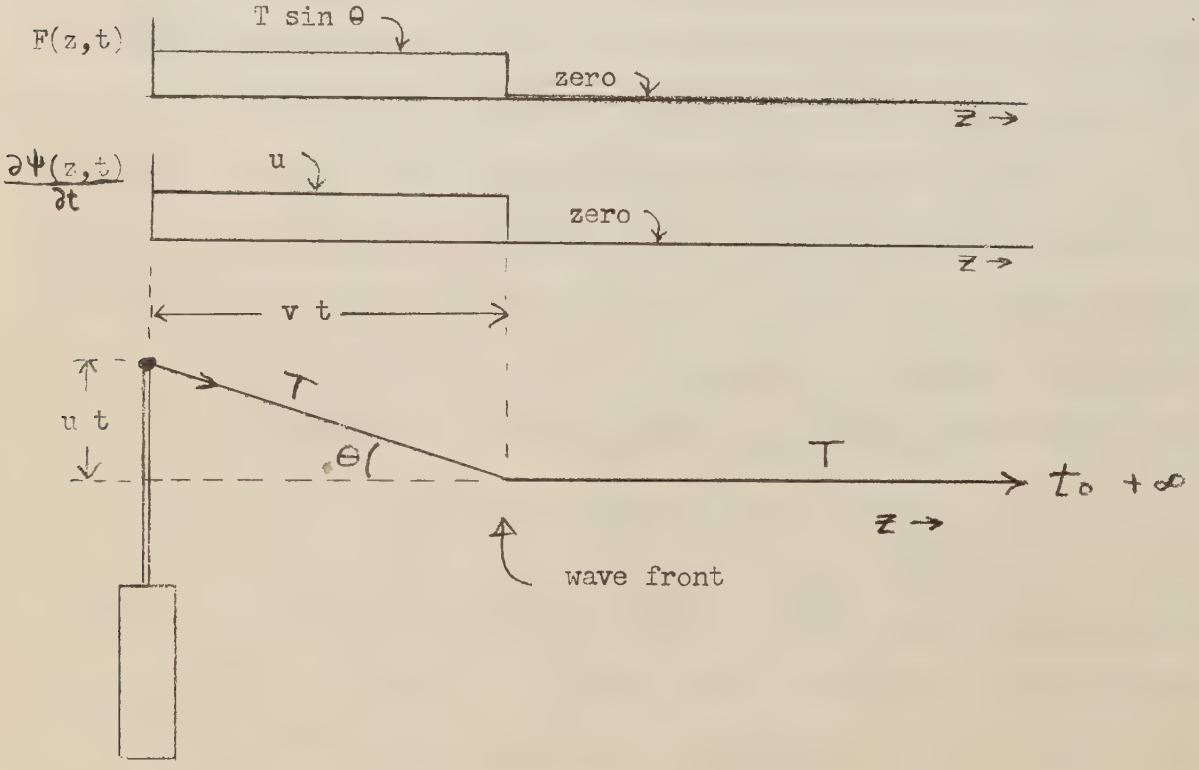


Fig. 3-2. Traveling-wave emitted by transmitter at $z = 0$. The transmitter output point is moving with velocity $u(t)$. The wave is propagating with velocity v. The motion of the transmitter started abruptly at $t = 0$. The function $F(z,t)$ is the transverse force exerted at time t on the portion of the string to the right of point z by the portion of the string to the left of point z. The function $\partial\psi(z,t)/\partial t$ is the transverse velocity of the string at z,t.

The transverse force $F(t)$ exerted by the transmitter output hook against the end of the string at $z = 0$ is shown positive in Fig. 2 (i.e., it is upwards). By Newton's third law this force is the negative of the downwards force exerted on the output hook by the transverse component of the string tension T. Thus

$$F(t) = T \sin\theta \approx T \tan\theta = T\frac{u(t)}{v} \quad , \tag{12}$$

where the last equality is obtained by using $\tan\theta = u/v$, a relation which is apparent from the figure. According to Newton's second law, the force $F(t)$ is equal to dp/dt, the rate of change of momentum imparted to the string (in the transverse direction). As we see in Fig. 2, the only place that the string does not have constant velocity [equal either to zero or to $u(t)$] is at the "wavefront." At the wavefront the amount of mass that acquires velocity increment u per unit time is ρv. Thus dp/dt is ρvu :

$$F(t) = \frac{dp}{dt} = \rho v u(t) \quad . \tag{13}$$

Combining Eqs. (12) and (13) gives

$$v = (T/\rho)^{\frac{1}{2}} \quad . \tag{14}$$

Note that this is constant, i.e., the string is non-dispersive. Defining the characteristic impedance Z experienced by the driving force $F(t)$ in accordance with Eq. (3), we obtain

$$Z = \rho v = T/v = (T\rho)^{\frac{1}{2}} \quad . \tag{15}$$

Energy considerations (including those for the general case). The power emitted by the transmitter (emitting a traveling wave) is

$$P = F(t)u(t) = \rho v u^2(t)$$

$$= \rho v \left[\frac{\partial \psi(0,t)}{\partial t}\right]^2 , \qquad (16)$$

where we have made use of the boundary condition at the transmitter,

$$u(t) = \frac{\partial \psi(0,t)}{\partial t} . \qquad (17)$$

Since the string is non-dispersive, i.e., v is constant, the group velocity is equal to the phase velocity v, as we saw in Chap. 2. Thus energy is propagated with the wave velocity v. Consequently, for a pure traveling wave the energy flux $S_z(z,t)$ (i.e., the energy per unit time passing point z at time t) should be the product of the energy density (energy per unit distance) times the propagation velocity v. (Dimensionally, we have energy/sec equals energy/cm times cm/sec.) By energy conservation the energy flux at point z at time t should equal the power emitted at $z = 0$ at the earlier time $t - (z/v)$. Thus by writing the energy flux at $z = 0$ [Eq. (16)] for a time $t - (z/v)$ and recalling the fact that, for a pure traveling wave, $\psi(z,t) = \psi[0, t - (z/v)]$, we can find the energy flux at (z,t) :

$$\text{flux } (z,t) = \text{flux } [0, t - (z/v)] = \rho v \left[\frac{\partial \psi[0, t - (z/v)]}{\partial t}\right]^2$$

$$= \rho v \left[\frac{\partial \psi(z,t)}{\partial t}\right]^2 . \qquad (18)$$

Consequently the total energy density at point z at time t is

$$\text{T.E.D.} = \begin{pmatrix}\text{total energy density}\\ \text{in traveling wave}\end{pmatrix} = \rho\left[\frac{\partial\psi(z,t)}{\partial t}\right]^2 . \qquad (19)$$

Now it is easy to see that the kinetic energy density is given in the general case, i.e., for standing and/or traveling waves, by

$$\text{K.E.D.} = \begin{pmatrix}\text{kinetic energy density}\\ \text{for general case}\end{pmatrix} = \tfrac{1}{2}\rho\left[\frac{\partial\psi(z,t)}{\partial t}\right]^2 . \qquad (20)$$

Thus we see that, according to Eqs. (19) and (20), the total energy density in a pure traveling wave is twice the kinetic energy density. The potential and kinetic energy densities are equal at all times and places for a pure traveling wave. See Fig. 3.

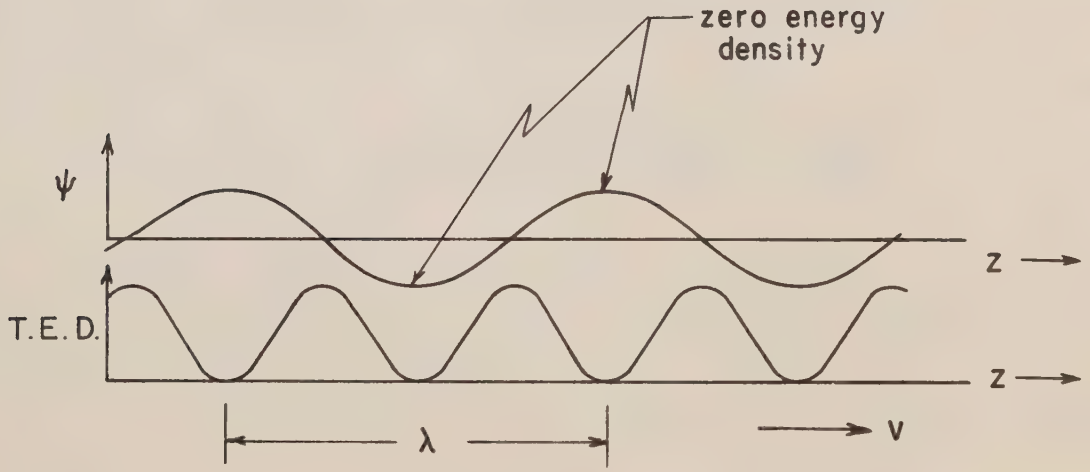


Fig. 3-3. Energy relations in harmonic traveling wave. The string displacement ψ and the total energy density (T.E.D.) are plotted versus z at a given instant t. The kinetic and potential energy densities are each half of the energy density and are not plotted. The entire pattern moves with velocity v. The places where the magnitude of the displacement ψ is maximum have zero kinetic energy density, because the string segments are momentarily at rest. These same places have zero potential energy density, because the string segments have zero slope and are therefore not stretched beyond their equilibrium lengths. At intermediate places where ψ is zero, the transverse string velocity is maximum, and the stretching due to the slope is also maximum. Therefore both kinetic and potential energy densities are maximum there.

We shall let you show (problem 1) that the expression for the potential energy density on a stretched string is given in the general case by

$$\text{P.E.D.} = \begin{pmatrix}\text{potential energy density} \\ \text{for general case}\end{pmatrix} = \tfrac{1}{2} T \left[\frac{\partial \psi(z,t)}{\partial z}\right]^2 . \qquad (21)$$

Suppose that instead of a pure traveling wave $F(z - vt)$ propagating in the $+z$ direction, we have a general superposition of waves traveling in both directions (which can describe standing waves as well -- see Sec. 2.5) given by

$$\psi(z,t) = F(z - vt) + G(z + vt) . \qquad (22)$$

Then we have

$$\frac{\partial \psi(z,t)}{\partial t} = - v \frac{dF}{d(z - vt)} + v \frac{dG}{d(z + vt)} , \qquad (23)$$

and

$$\frac{\partial \psi(z,t)}{\partial z} = \frac{dF}{d(z - vt)} + \frac{dG}{d(z + vt)} . \qquad (24)$$

Then, letting $dF/d(z - vt) \equiv F'$ and $dG/d(z + vt) \equiv G'$ (and using $T = \rho v^2$), we see that the kinetic and potential energy densities in the general case are [using Eqs. (20), (21), (23), and (24)]

$$\text{K.E. density} = \tfrac{1}{2}\rho v^2 \left\{F'^2 + G'^2 - 2F'G'\right\} \qquad (25)$$

$$\text{P.E. density} = \tfrac{1}{2}\rho v^2 \left\{F'^2 + G'^2 + 2F'G'\right\} . \qquad (26)$$

Thus in the general case where both $F(z - vt)$ and $G(z + vt)$ are present, the kinetic and potential energy densities are not equal, because the "interference term" $2F'G'$ enters with opposite signs for kinetic and potential energy densities. There are two interesting cases where the K.E. and P.E. densities are equal. One is the case where you have a pure traveling wave,

such that you have $F(z - vt)$ alone (or $G(z + vt)$ alone). Then $F'G'$ vanishes. The other case of interest is when you have two pulses traveling in opposite directions and far enough apart so that they do not "overlap," i.e., at those z where F is non-zero, G is zero, and vice versa.

Notice that the phrase "pure traveling wave" does not necessarily mean a harmonic traveling wave but can mean any superposition of harmonic traveling waves all of which travel in the same direction, i.e., the general "pure" traveling wave $F(z - vt)$ [or $G(z + vt)$] .

Flux vector. One can define a flux vector $\underset{\sim}{S}(z,t)$ which is directed along $\pm z$ and whose z component is the rate of flow of energy in the $+z$ direction at time t and position z . The flux vector is given by the general expression

$$S_z(z,t) = \left[- T \frac{\partial \psi(z,t)}{\partial z} \right] \frac{\partial \psi(z,t)}{\partial t} , \qquad (27)$$

which we derive as follows: The transverse force exerted on the portion of the string just to the right of z by that portion just to the left is given (by inspection of Fig. 2, and using the point $z = 0$) by

$$F(\text{L on R}) = - T \frac{\partial \psi(z,t)}{\partial z} . \qquad (28)$$

(For Fig. 2, $\partial\psi/\partial z$ is negative and is $- \tan \theta \approx - \sin \theta$) . Equation (28) is completely general, i.e., it holds for any superposition of waves. Now, the transverse velocity of the string at (z,t) is $\partial\psi/\partial t$. The product of the force exerted by the string to the left of z on that to the right times the velocity of that point on the string equals the rate at which the string to the left of z is doing work on the string to the right of z . This product is therefore the rate of flow of energy in the $+z$ direction at (z,t) . Thus the flux S_z is indeed given by Eq. (27). For a traveling

wave traveling in the $+z$ direction, $\partial\psi/\partial z$ is $-v^{-1}\,\partial\psi/\partial t$, as can be seen by differentiating $F(z - vt)$. In that case S_z is $+(T/v)(\partial\psi/\partial t)^2$, according to Eq. (27). Since T is ρv^2 and S_z is the energy density times v for a traveling wave, this agrees with our previous result, Eq. (19).

Energy density in pure standing waves. The energy relations in a standing wave are quite different from those in a traveling wave. Consider the standing wave

$$\psi(z,t) = A \sin kz \cos \omega t = A \sin kz \cos vk\, t \qquad (29)$$

(we used $\omega = vk$ in the second equality). Using Eqs. (20) and (21) with Eq. (29) we find

$$\text{K.E. density} = \tfrac{1}{2} A^2 \rho\, v^2 k^2 \sin^2 kz \sin^2 \omega t\ , \qquad (30)$$

$$\text{P.E. density} = \tfrac{1}{2} A^2 \rho\, v^2 k^2 \cos^2 kz \cos^2 \omega t\ . \qquad (31)$$

Thus at time zero the P.E. density is maximum at $z = 0$ and at $\lambda/2$ and is zero halfway in between, i.e., at $z = \lambda/4$. At the same time (zero) the kinetic energy density is zero everywhere. One quarter cycle later, at $t = \frac{1}{4} T$ (T is the period, not the tension), the potential energy density is zero everywhere, and the kinetic energy density is maximum at $z = \lambda/4$ and is zero at $z = 0$ and $\lambda/2$. This behavior implies that energy flows back and forth from one place to another, changing from kinetic energy at one place to potential energy at another. Thus S_z alternates in sign. These relations are shown in Fig. 4.

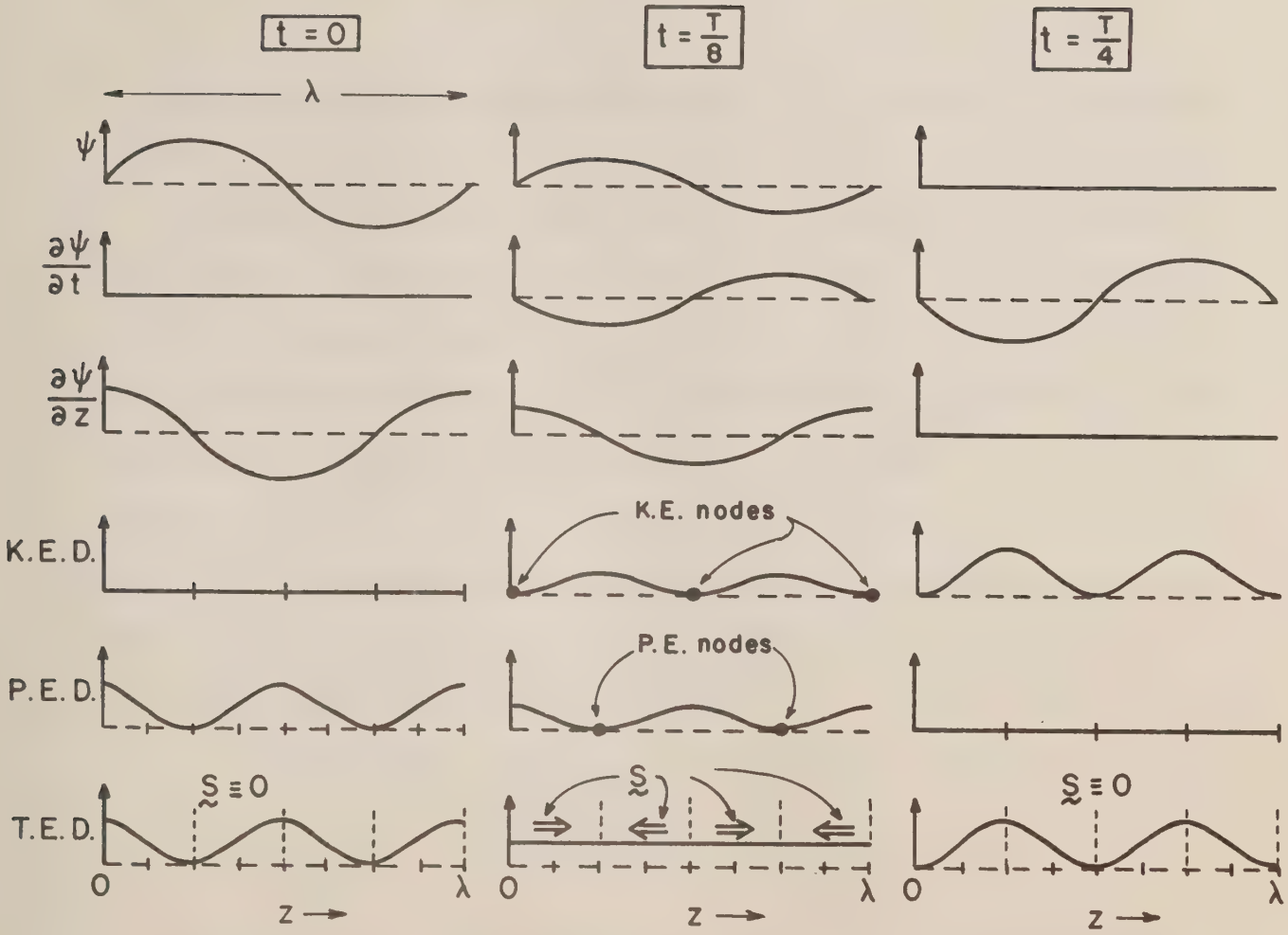


Fig. 2-4. Harmonic standing wave on a string. The string displacement ψ, velocity $\partial\psi/\partial t$, and slope $\partial\psi/\partial z$, and the kinetic, potential, and total energy densities (K.E.D., P.E.D., and T.E.D.) are plotted over a range of one wavelength for times $t = 0$, $T/8$, and $T/4$. The energy flux vector $\underset{\sim}{S}$ is zero for all z at $t = 0$ when the energy is entirely potential, and at $t = T/4$ when the energy is entirely kinetic. At $t = T/8$ the flux $\underset{\sim}{S}$ is greatest at the places indicated by double-shafted arrows in the plot of T.E.D. versus z. The flux $\underset{\sim}{S}$ is always zero at every node (zero of ψ) and antinode (maximum of ψ). Note that the energy flux at $t = T/8$ is such as to change the distribution of T.E.D. from that at $t = 0$ to that at $t = T/4$.

Reflection of an incident traveling wave at a mismatched resistive load. Suppose the string is terminated in a "dashpot," as in Fig. 1. A dashpot is a device with a moveable part to which the string can be connected (terminated), and which has the property that the moveable part responds to being driven at velocity u with a "resistive drag" response force equal to $-Z_2 u$, where Z_2 is called the impedance, or "load impedance," of the dashpot. If Z_2 equals Z_1, the characteristic impedance of the string, the load is said to match impedances with the medium (the string), and the incident traveling wave is then absorbed without reflection.

Let us see what happens when Z_1 and Z_2 are unequal. Let the string extend from z equals $-\infty$ to z equals zero, with the dashpot located at $z = 0$. An incident traveling wave ψ_{inc} comes in from the left:

$$\psi_{inc} = A \cos(\omega t - kz) \quad . \tag{32}$$

A reflected wave ψ_{ref} is generated at the load at $z = 0$, where the impedance suddenly changes from Z_1 to the load impedance Z_2. The reflected wave is given, we assume, by

$$\psi_{ref} = R_{12} A \cos(\omega t + kz) \quad , \tag{33}$$

where R_{12} is an unknown reflection coefficient which we wish to determine. The complete wave function $\psi(z,t)$ is given by the superposition of the incident and reflected waves:

$$\psi(z,t) = \psi_{inc} + \psi_{ref} = A \cos(\omega t - kz) + R_{12} A \cos(\omega t + kz). \tag{34}$$

Now we must consider the boundary conditions at $z = 0$. According to our derivation of Eq. (28), the string at $z = 0$ exerts a force on the load

given by

$$F \text{ (string on load)} = - T \left[\frac{\partial \psi(z,t)}{\partial z} \right]_{z=0} . \qquad (35)$$

The force exerted by the load on the string is the drag force

$$F \text{ (load on string)} = - Z_2 \left[\frac{\partial \psi(z,t)}{\partial t} \right]_{z=0} . \qquad (36)$$

But according to Newton's third law, the sum of Eqs. (35) and (36) vanishes:

$$T \frac{\partial \psi(z,t)}{\partial z} + Z_2 \frac{\partial \psi(x,t)}{\partial t} = 0 , \text{ at } z = 0 . \qquad (37)$$

Inserting the wave function ψ as given by Eq. (34) into the boundary condition Eq. (31) [and remembering to <u>first</u> take the partial derivative with respect to z, and <u>then</u> set $z = 0$] we find

$$T k \sin \omega t - T k R_{12} \sin \omega t - Z_2 \omega \sin \omega t - Z_2 \omega R_{12} \sin \omega t = 0 .$$

This gives (setting k equal to ω/v)

$$R_{12} = \frac{(T/v) - Z_2}{(T/v) + Z_2} . \qquad (38)$$

According to Eq. (15) this is equivalent to

$$\boxed{R_{12} = \frac{Z_1 - Z_2}{Z_1 + Z_2}} . \qquad (39)$$

We are not really interested in strings. When we study sound waves we shall sometimes be interested in the velocity of the air and sometimes in the pressure. It turns out that the velocity behaves in a manner analogous to $\partial\psi/\partial t$ for the string and that the pressure behaves analogously to $\partial\psi/\partial z$ for the string. We are therefore interested not only in the reflection coefficient for $\psi(z,t)$, but also in those for $\partial\psi/\partial t$ and $\partial\psi/\partial z$. We obtain

these reflection coefficients by differentiating Eq. (34) with respect to t and with respect to z :

$$\frac{\partial\psi(z,t)}{\partial t} = -\omega\,[A \sin(\omega t - kz) + R_{12} A \sin(\omega t + kz)] \tag{40}$$

$$\frac{\partial\psi(z,t)}{\partial z} = k\,[A \sin(\omega t - kz) - R_{12} A \sin(\omega t + kz)] \ . \tag{41}$$

Thus we see that $\partial\psi/\partial t$ reflects with a reflection coefficient R_{12} which is the same as that of ψ . However, $\partial\psi/\partial z$ reflects with a coefficient $-R_{12}$ which is the negative of that of $\partial\psi/\partial t$ and of ψ :

$$R_{\psi} = R_{\partial\psi/\partial t} = R_{12} \tag{42}$$

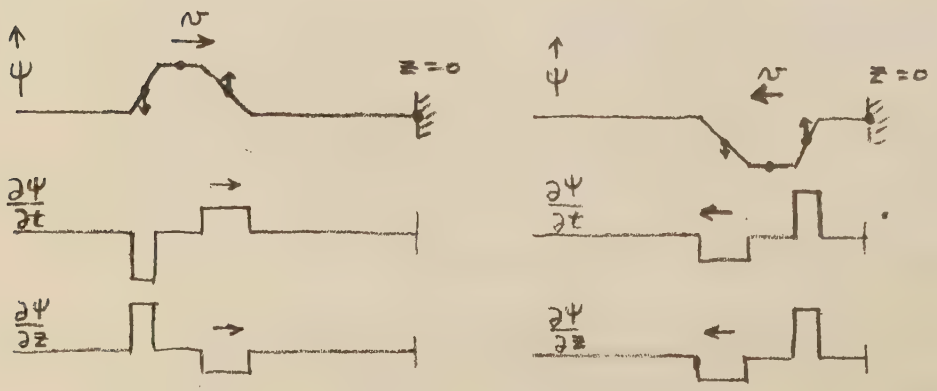
$$R_{\partial\psi/\partial z} = -R_{12} \ , \tag{43}$$

where R_{12} is given by Eq. (39).

Here are some interesting limiting cases:

1. Perfect termination. (See Fig. 1.) If Z_2 equals Z_1 there is no reflected wave.

2. Infinite drag -- fixed end point. (See Fig. 5.) (Note that since we are now considering a non-dispersive string, we have the same behavior for a pulse as for a harmonic wave. Therefore we are free to illustrate using a pulse, as in Fig. 5.)

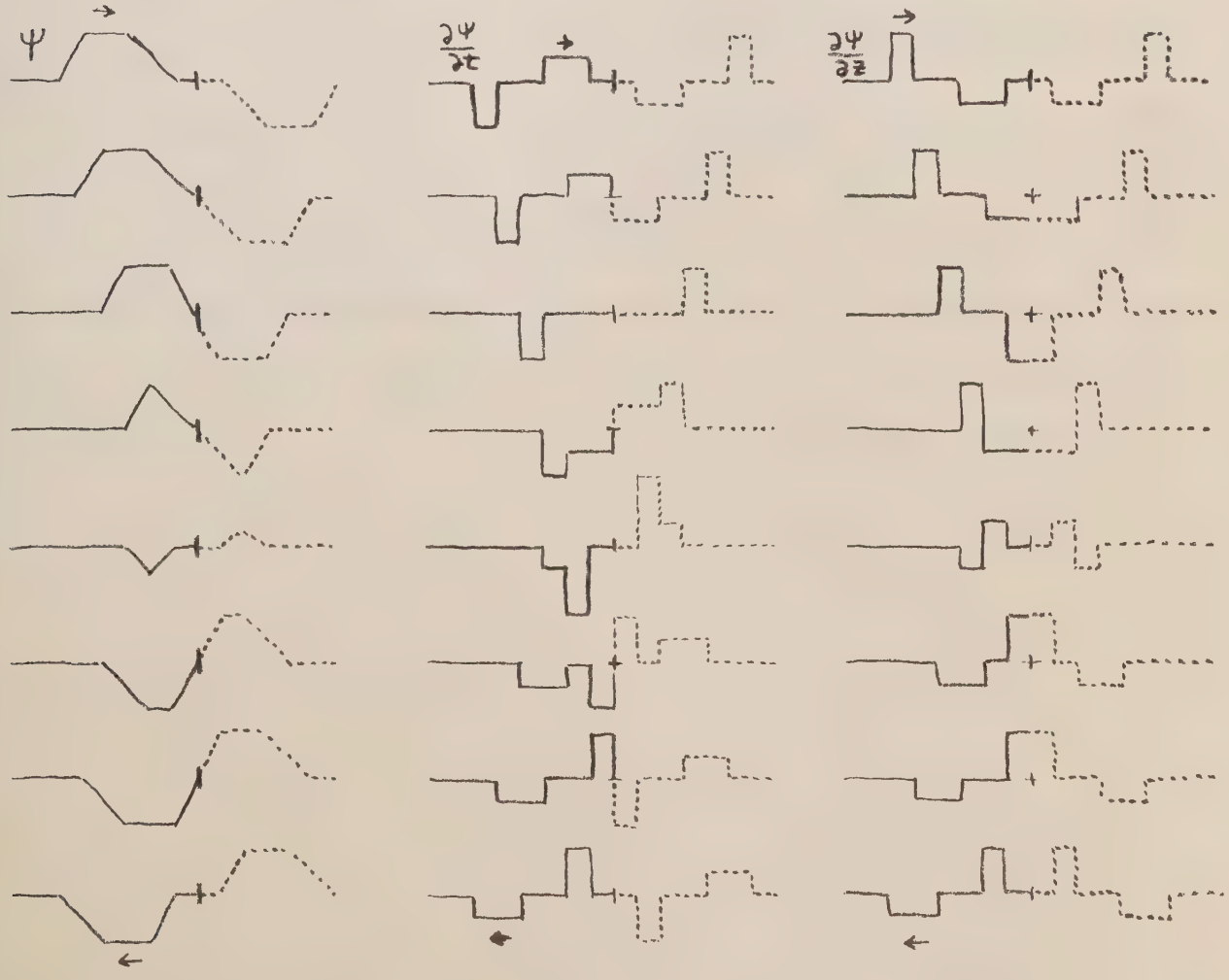


(a) Before reflection (b) After reflection

Fig. 3-5. Reflection of an incident pulse at a fixed end of the string. The small vertical arrows on the three points indicate the instantaneous velocity of the string at those points. (The center one is a zero-length arrow.)

If Z_2 is infinite the dashpot terminal is immovable and can therefore absorb no energy. The reflection coefficients R_ψ and $R_{\partial\psi/\partial t}$ are -1, so that ψ and $\partial\psi/\partial t$ reverse sign upon reflection. The reflection coefficient $R_{\partial\psi/\partial z}$ is +1, and thus $\partial\psi/\partial z$ has the same sign after reflection as before.

During the reflection process, when both the incident and reflected traveling waves are present at $z = 0$, ψ and $\partial\psi/\partial t$ at $z = 0$ are required to remain zero. The slope $\partial\psi/\partial z$ at $z = 0$ is twice as great as it would be if there were perfect termination. In Fig. 6 we examine in detail the reflection of the pulse shown in Fig. 5. (It may help you to think of the reflected pulse as existing "virtually" to the right of $z = 0$, i.e., "inside the wall,"



(a) Displacement (b) Velocity (c) Slope

Fig. 3-6. Reflection at fixed end of string. The dotted curve is the virtual reflected pulse "before the reflection occurs."

and reaching z = 0 at just the right time to help satisfy the boundary condition that the string is fixed to the wall at z = 0. If there were a real continuation of the string to the right of z = 0, instead of a virtual string, and if it carried a real pulse that looked like the imagined virtual pulse, then you could eliminate the wall (the infinitely draggy dashpot) and the point at z = 0 would still remain motionless.)

If the incident wave is a harmonic wave, then for a fixed end we get [using $R_{12} = -1$ in Eqs. (34), (40), and (41)]

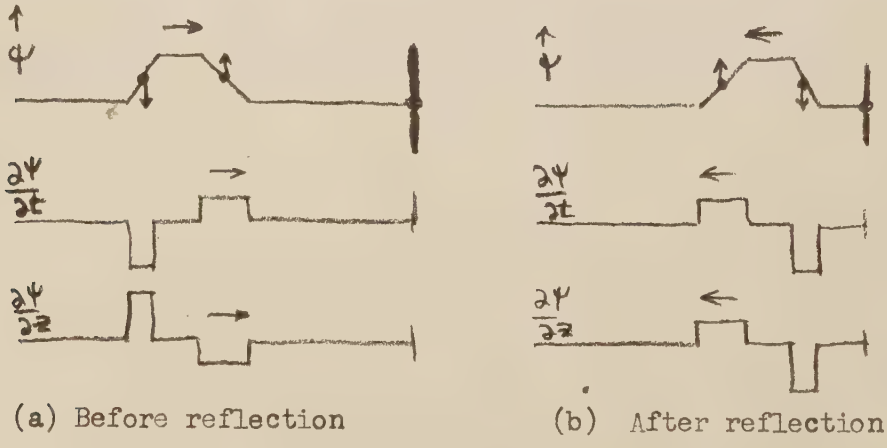
$$\psi(z,t) = A \cos(\omega t - kz) - A \cos(\omega t + kz) = 2A \sin \omega t \sin kz \tag{44}$$

$$\frac{\partial \psi(z,t)}{\partial t} = -\omega A[\sin(\omega t - kz) - \sin(\omega t + kz)] = +2\omega A \cos \omega t \sin kz \tag{45}$$

$$\frac{\partial \psi(z,t)}{\partial z} = kA[\sin(\omega t - kz) + \sin(\omega t + kz)] = 2kA \sin \omega t \cos kz \tag{46}$$

Thus ψ and $\partial\psi/\partial t$ are standing waves with a node (zero) at $z = 0$, and $\partial\psi/\partial z$ is a standing wave with an antinode (maximum) at $z = 0$.

3. Zero drag. Free end. See Fig. 7.



(a) Before reflection (b) After reflection

Fig. 3-7. Reflection of a pulse from a free end of a string. (The string is attached to a massless ring which slides on a frictionless rod.) The small vertical arrows give velocities of the string at those points.

If Z_2 is zero the dashpot offers no resistance. The string can therefore exert no force on the dashpot. (According to both Newton's third law and common experience, you cannot push on something that can't push back.) Since the transverse force exerted by the string is proportional to the slope, the slope must remain zero at a free end. No energy is absorbed by the dashpot (because there is no force on it). The transverse velocity and displacement reflect with a reflection coefficient of +1 and have values twice what they would be for perfect termination. In Fig. 8 we follow ψ through the reflection.

For an incident harmonic wave, reflection at a free end gives standing waves, as you can see by substituting $R_{12} = +1$ in Eqs. (34), (40), and (41). In this case, the displacement and velocity have antinodes at $z = 0$, and the slope has a node there.

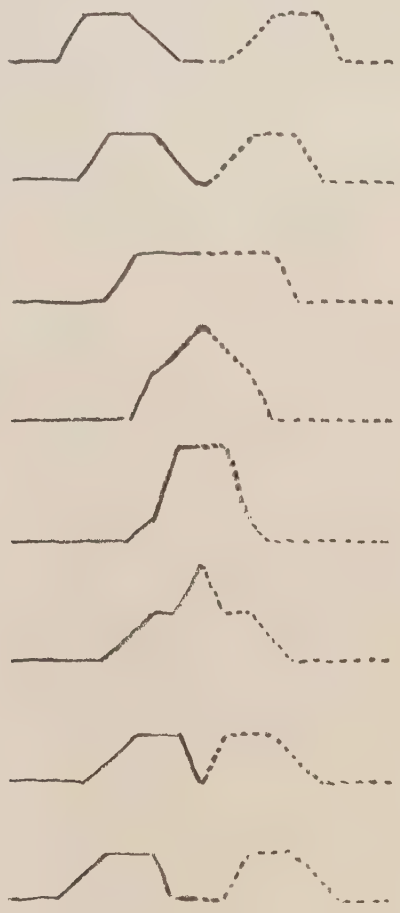


Fig. 3-8. Reflection at a free end. The dashed pulse is a virtual wave representing the reflected traveling wave "before the reflection occurs." Superposition of the incident and reflected pulses gives the total displacement for $z \leq 0$. The slope remains zero at $z = 0$. (It doesn't appear to do so in some of the frames. That is because we drew oversimplified pulses having discontinuities in slope.)

3.3 Sound waves.

We shall first consider a one-dimensional system consisting of an infinitely long tube filled with air which is driven by a piston that can move longitudinally along the tube. The piston is the "moving part" that couples the transmitter to the medium. At equilibrium, the piston is at $z = 0$ with zero velocity, and the gas has its equilibrium pressure p_o and equilibrium density ρ_o. See Fig. 9.

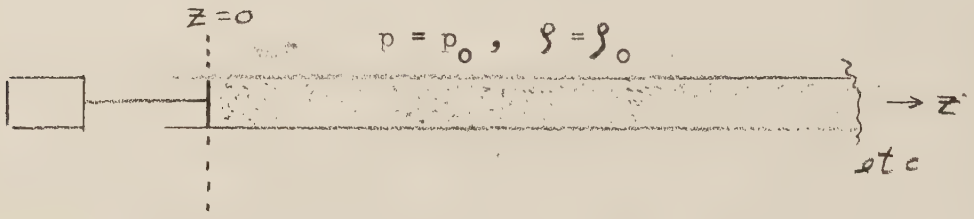


Fig. 3-9. Gas-filled tube driven by piston. Equilibrium conditions.

When the piston is in motion, it has displacement $D(t)$ from its equilibrium position at $z = 0$. The displacement $D(t)$ is positive if it is in the $+z$ direction. The velocity of the piston is $u(t)$:

$$\text{piston velocity} = u(t) = \frac{d\,D(t)}{dt} \quad . \tag{47}$$

We think of the gas as a continuous compressible fluid. Sometimes it helps to picture the individual molecules, but we are actually ignoring all effects due to the "graininess" resulting from the molecular character of the gas. For instance, when we speak of "the displacement of the molecules whose equilibrium position is z " we mean the <u>average</u> displacement of a huge number of molecules in a tiny volume element located at z . (For air

at STP a cube of edge 10^{-5} cm contains about 30,000 molecules.) Thus we can ignore the rapid and erratic thermal motions of the individual gas molecules, because these average to zero when we average over a tiny volume element. For convenience and simplicity we shall think of the molecules as having no motion except the average motion, so that neighboring molecules remain neighbors. This simplifies our mental picture and introduces no errors into our predictions. (Sound as a wave motion in a continuous medium was correctly understood long before molecules were discovered.) To enhance this picture we shall often call the gas a "fluid."

Description of the gas. Now we make up some notation to describe the average motion of the gas. This motion is induced by the longitudinal motion of the piston. Therefore the gas motion is longitudinal, back and forth along the z axis. We let $\psi(z,t)$ represent the instantaneous displacement in the $+z$ direction of a small amount of gas whose equilibrium position is z. Thus when $\psi(z,t)$ is zero the gas is at z. When $\psi(z,t)$ is not zero the gas is at position $z + \psi(z,t)$. It is important to remember that in all the quantities we shall use (displacement, velocity, pressure, density) which are expressed as a function of z and t, z always means the equilibrium position. [It is very easy to become confused, because of the fact that the longitudinal motion of the vibrating gas is along the propagation direction z of the sound waves. Thus, for example, when someone says "the mass density is a function of position and time, $\rho(z,t)$," you might wonder whether z means the place where the gas is now (at time t) or where it was (at equilibrium). By $\rho(z,t)$ we will mean the density at time t in the tiny volume element whose equilibrium position is at z, wherever that volume element happens to be at time t.]

Thus

$$\left\{\begin{array}{c}\text{instantaneous position}\\ \text{of that fluid}\\ \text{whose equilibrium position is } z\end{array}\right\} = z + \psi(z,t) \quad . \tag{48}$$

(Note that going to new z always means considering new molecules, i.e., a new volume element of fluid.)

The instantaneous velocity of the fluid is the rate of change of position with time, i.e., according to Eq. (48) the partial derivative of ψ with respect to t, with z held constant. (Holding z constant just means that we consider a given "molecule," by which we mean a suitable "average molecule" which has the average motion of a small amount of fluid that is centered at z at equilibrium.) We shall call this velocity by the name $v(z,t)$:

$$\left\{\begin{array}{c}\text{instantaneous velocity}\\ \text{of fluid}\\ \text{whose equilibrium position is } z\end{array}\right\} = v(z,t) = \frac{\partial\psi(z,t)}{\partial t} \quad . \tag{49}$$

Two other quantities of importance are the instantaneous density $\rho(z,t)$ and the instantaneous pressure $p(z,t)$. These two quantities are closely related. When the density increases, so does the pressure. We shall look at the relation between pressure and density presently. First we wish to find the relation between $\rho(z,t)$ and $\psi(z,t)$. We shall write down the answer; then we shall derive it:

$$\left\{\begin{array}{c}\text{instantaneous density}\\ \text{of fluid}\\ \text{whose equilibrium position is } z\end{array}\right\} = \rho(z,t) = \rho_o - \rho_o \frac{\partial\psi(z,t)}{\partial z} \quad . \tag{50}$$

To obtain Eq. (50), consider Fig. 10.

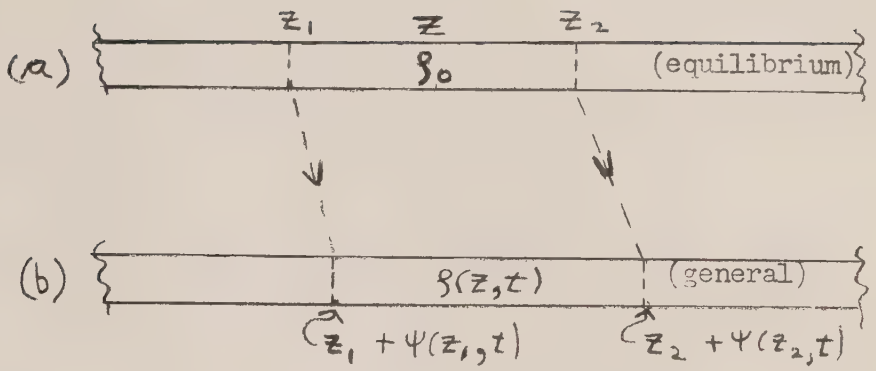


Fig. 3-10. Relation between density and displacement.

At equilibrium [Fig. 10(a)] the total amount of mass between z_1 and z_2 is $\rho_o\ (z_2 - z_1)$ times the cross sectional area of the tube, which we take to be one square centimeter. When we are not at equilibrium [Fig. 10(b)], those molecules originally at z_1 are at $z_1 + \psi(z_1,t)$. Those originally at z_2 are at $z_2 + \psi(z_2,t)$. (If you want to worry about the random motions we can amend that last remark to read "the same number of molecules, on the average." We don't care whether they are the very same molecules. One molecule is as good as another.)

Let us define an <u>average equilibrium position</u> z (lying approximately half way between z_1 and z_2) such that $\rho(z,t)$ [the density at time t of the fluid element which has equilibrium position z] equals the average density at time t between $z_1 + \psi(z_1,t)$ and $z_2 + \psi(z_2,t)$. Using this definition, we have by conservation of mass that

$$\rho_o(z_2 - z_1) = \rho(z,t)[z_2 + \psi(z_2,t) - z_1 - \psi(z_1,t)] \quad . \tag{51}$$

Now we let z_2 approach z_1 , until it is so close that we can neglect quadratic and higher powers of $(z_2 - z_1)$ in the Taylor's series expansion

of $\psi(z_2,t)$ at (z_1,t) :

$$\psi(z_2,t) = \psi(z_1,t) + (z_2-z_1)\left[\frac{\partial\psi(z,t)}{\partial z}\right]_{z_1} + \tfrac{1}{2}(z_2-z_1)^2\left[\frac{\partial^2\psi(z,t)}{\partial z^2}\right]_{z_1} + \ldots \quad (52)$$

$$\approx \psi(z_1,t) + (z_2-z_1)\left[\frac{\partial\psi(z,t)}{\partial z}\right]_{z} \quad (53)$$

Inserting Eq. (53) into Eq. (51) we obtain

$$\rho(z,t) = \rho_o\left\{1 + \frac{\psi(z_2,t) - \psi(z_1,t)}{z_2 - z_1}\right\}^{-1} \quad (54)$$

$$\approx \rho_o\left[1 + \frac{\partial\psi(z,t)}{\partial z}\right]^{-1} \quad (55)$$

$$= \rho_o\left\{1 - \frac{\partial\psi(z,t)}{\partial z} + \left[\frac{\partial\psi(z,t)}{\partial z}\right]^2 + \ldots\right\} \quad (56)$$

$$\approx \rho_o - \rho_o\frac{\partial\psi(z,t)}{\partial z}\ , \quad (57)$$

which is Eq. (50). [Equation (54) is Eq. (51) rearranged; Eq. (55) is obtained after substituting Eq. (53); Eq. (56) follows from Eq. (55) and the series $(1 + x)^{-1} = 1 - x + x^2 - \ldots$, given in the Appendix, Eq. (20), with $n = -1$ and $x = \partial\psi/\partial z$; Eq. (57) follows from Eq. (56) when we neglect the non-linear terms. It turns out that neglecting non-linear terms gives a very good approximation for ordinary sound. In intense shock waves from explosions the non-linear terms may be important.] We have given you all the details for clarity. But when you "practice" deriving Eq. (50) you should drop the higher order terms without ever writing them down. The physicist's approach is usually to retain only the first terms in any theory

(which might be wrong, even in first order), until the experiments improve enough to justify the mental labor of looking at higher order contributions.

Our final quantity of interest is the pressure $p(z,t)$. At equilibrium the pressure is uniform and equal to p_o. The pressure p can be expressed as a function of the mass density ρ and the absolute temperature T. (You will not officially study temperature and the behavior of gases until Vol. 5. If you feel unsatisfied with what you find here, read "The Nature of a Gas," in that great high school book "Physics," by PSSC, Chap. 9, 2nd edition, D. C. Heath and Company, Boston.) At equilibrium the temperature is T_o. When the gas is in motion, the temperature may vary. In general we expect $T = T(z,t)$, where, as usual, z refers to the equilibrium position of the little bit of fluid under consideration, and $T(z,t)$ is the "local temperature" of the displaced fluid. The reason that we do not expect T to remain at the constant value T_o is that when one part of the gas gets compressed (by being pushed together by its neighbors), work is done in compressing the gas, so that it tends to heat up above T_o when compressed above density ρ_o. Conversely, the places with ρ less than ρ_o are cooler than T_o. We can write the relation between pressure p, density ρ, and temperature T by considering only the first terms in a Taylor's series expansion of $p(\rho,T)$ around the values ρ_o, T_o:

$$p = p_o + (\rho - \rho_o)\left(\frac{\partial p}{\partial \rho}\right)_o + (T - T_o)\left(\frac{\partial p}{\partial T}\right)_o , \qquad (58)$$

where the subscript zero means that the derivative is to be evaluated at the equilibrium pressure and temperature. According to the "perfect gas law," which gives a good description of air at STP, the relation between p, ρ, and T is

$$p = p_o \frac{\rho}{\rho_o} \frac{T}{T_o} , \tag{59}$$

so that

$$\left(\frac{\partial p}{\partial \rho}\right)_o = \frac{p_o}{\rho_o} , \quad \left(\frac{\partial p}{\partial T}\right)_o = \frac{p_o}{T_o} . \tag{60}$$

Then Eq. (58) becomes

$$p = p_o + (\rho - \rho_o) \frac{p_o}{\rho_o} + (T - T_o) \frac{p_o}{T_o} . \tag{61}$$

We still lack the relation between pressure and density until we know the relation between the temperature increase $T - T_o$ and the density increase $\rho - \rho_o$. We could assume that the processes in a sound wave are isothermal (i.e., at constant temperature) and set $T-T_o = 0$, but that would be wrong. (This is what Isaac Newton assumed. It gives the wrong answer for the sound velocity in air. It predicts a velocity of 280 m/sec at STP, whereas the observed value is 332 m/sec.)

We will assume instead that the processes in the gas are "adiabatic." This just means that no energy flows in the form of heat from volume elements which are denser than ρ_o and consequently hotter than T_o ("crests" of the pressure wave) to elements which are less dense than ρ_o and consequently colder than T_o ("troughs" of the pressure wave). To justify this assumption fully would require an excursion into statistical mechanics, but the crucial point is as follows: The thermal energy in a gas appears as random motions of the molecules all bouncing against one another. If the molecules with high thermal energy (those in the crests) could bounce on those with low thermal energy (those in the troughs), thermal energy would be transferred from hot to cold regions, and our assumption would be invalid. It turns out that the average thermal velocity is about the same as the velocity of sound waves

(since this is about how fast a molecule struck by an advancing wave front will transmit the "signal" to those ahead.) However, the thermal motions zig-zag in random directions, so the average distance covered by a "hot" molecule per half cycle of the wave is small compared to the distance covered by the wave itself in the same time, i.e., a half wavelength. Thus there is effectively no heat transfer and the adiabatic assumption is a good one.

It turns out that for an adiabatic process ρ and T are related by

$$T = T_o (\rho/\rho_o)^{\gamma-1} \quad , \tag{62}$$

where γ is a constant called "the ratio of specific heats" which is characteristic of the structure of the gas molecule. For an ideal gas of non-interacting point molecules [one satisfying Eq. (59)], γ is 1.667. For helium at STP γ is 1.660. For air at STP γ is 1.40. The reason that air does not have the ideal gas value, while helium does (approximately), is that the diatomic N_2 and O_2 molecules of air can tumble and vibrate as well as translate, whereas helium is monatomic and behaves somewhat like a smooth sphere. When air is compressed adiabatically, some of the energy goes into tumbling and vibrating. This slows the rate of rise of the translational energy, which is proportional to the temperature.

We ask you to accept Eq. (62) without proof. Then by differentiating it we obtain

$$\frac{dT}{T} = (\gamma-1)\frac{d\rho}{\rho} \quad , \tag{63}$$

or, for $T - T_o$ and $\rho - \rho_o$ small (so as to neglect higher terms in the Taylor's series, as usual),

$$\frac{T - T_o}{T_o} = (\gamma - 1)\frac{(\rho - \rho_o)}{\rho_o} \quad . \tag{64}$$

Inserting Eq. (64) into Eq. (61) we obtain the derired relation between pressure and density :

$$p = p_o + \frac{\gamma p_o}{\rho_o}(\rho - \rho_o) \quad . \tag{65}$$

Finally we combine Eqs. (65) and (50) to obtain

$$\left\{\begin{array}{c}\text{instantaneous pressure} \\ \text{of fluid} \\ \text{whose equilibrium position is } z\end{array}\right\} = p(z,t) = p_o - \gamma p_o \frac{\partial \psi(z,t)}{\partial z} \quad . \tag{66}$$

In Eqs. (48), (49), (50), and (66) we have the relations between the fluid displacement [given by $\psi(z,t)$], velocity [given by $\partial\psi(z,t)/\partial t$], density excess $\rho(z,t) - \rho_o$ [given by $-\rho_o \partial\psi(z,t)/\partial z$], and gauge pressure $p(x,t) - p_o$ [given by $-\gamma p_o \partial\psi(z,t)/\partial z$]. These are all interesting quantities, but we do not need to keep track of all of them. The velocity and the pressure are the most interesting, and they are sufficient to describe the waves. [That is, the two derivatives $\partial\psi(z,t)/\partial t$ and $\partial\psi(z,t)/\partial z$ are sufficient. For example, we can find $\psi(z,t)$ if we are given $\psi(0,t)$ and know $\partial\psi(z,t)/\partial t$ and $\partial\psi(z,t)/\partial z$ for all z and t. This will be shown later in this Section. Similarly we can find the density excess given the gauge pressure (and the equilibrium pressure and density).] The fluid velocity $v(z,t)$ and gauge pressure $p(z,t) - p_o$ have the desirable property that they are directly measureable, using suitable "flow meters" and pressure gauges. The displacement $\psi(z,t)$ is less easily measured. From now on we shall focus our attention on velocity and pressure.

Emission of traveling sound waves. We shall consider a very simple situation strictly analogous to that which we considered in Fig. 2 to derive the wave velocity and characteristic impedance for transverse waves on a string. We suppose that for $t < 0$ the system (shown in Fig. 9) is at equilibrium, with the piston resting at $z = 0$, the pressure $p(z,t)$ equal to p_o for all z, and the fluid velocity $v(z,t)$ equal to zero for all z. The force exerted on the piston by the drive rod is $p_o A$, where A is the area of the tube (and of the piston). If it weren't for this force the gas would push the piston out of the tube. We do not want to include the drive rod in our system. We shall merely say that at equilibrium the piston exerts pressure p_o (or force Ap_o) on the fluid in contact with the piston at $z = 0$, without worrying about how the piston does it.

At $t = 0$ the pressure exerted by the piston (on the fluid at its face) is suddenly increased from p_o to a new value, which we call p, and is then maintained at p. Just after $t = 0$, the fluid in a small layer next to the piston has pressure p exerted on it by the piston and pressure p_o by the adjacent fluid at greater z. The pressure differential on this fluid element therefore causes it to start to move at velocity u. In order to maintain the pressure p the piston must also move with velocity u. The first bit of fluid pushes on the next, raising its velocity from zero to u. That pushes on the next, etc. A "wave front" (the boundary where the pressure suddenly increases from p_o to p and the velocity from zero to u) propagates down the tube with the wave propagation velocity v (an unknown that we wish to find). This process is shown in Fig. 11. (It is interesting to compare Fig. 11 with the analogous one for the string, Fig. 2).

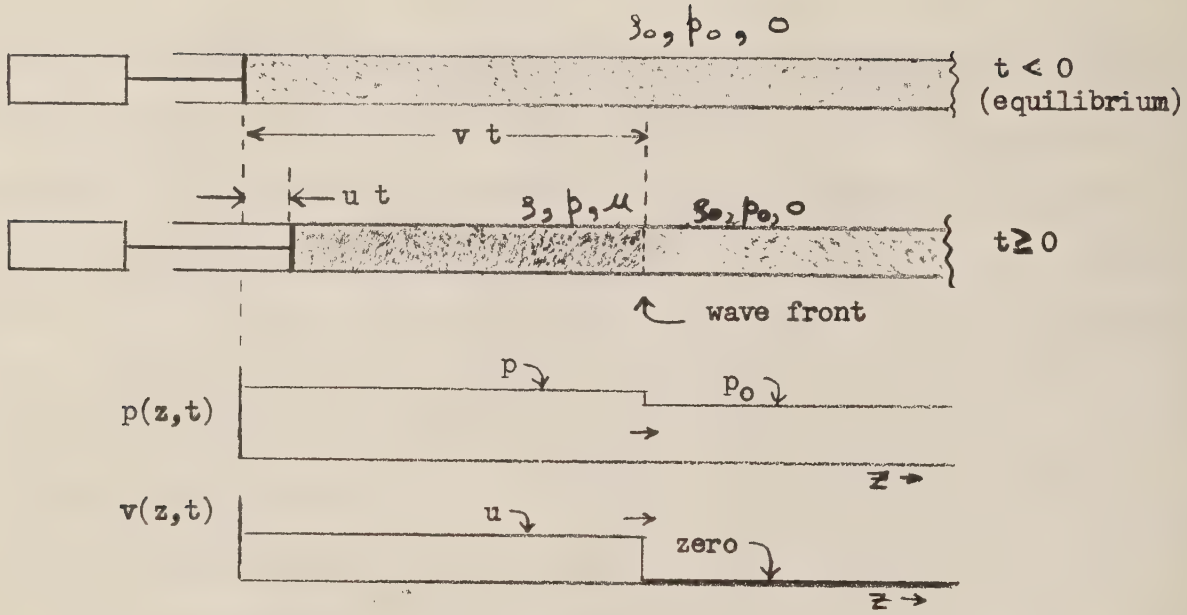


Fig. 3-11. Traveling sound wave in tube. For $t \geq 0$ the piston moves with steady velocity u. A wave front of compression travels along the tube with velocity v.

Notice that in the above description we have already implicitly given the boundary condition at the transmitter, which is that the fluid remains in contact with the piston:

$$\text{displacement:} \quad D(t) = \psi(0,t) \tag{67}$$

$$\text{velocity:} \quad u(t) = \frac{\partial \psi(0,t)}{\partial t} \ . \tag{68}$$

In our simple example, $u(t)$ is zero for $t < 0$ and has constant value u for $t \geq 0$. The velocity u is small compared with v in all cases we shall consider.

We can obtain the following "geometrical" relation from Fig. 11 (derived below):

$$\frac{u}{v} = \frac{\rho - \rho_0}{\rho_0} \ . \tag{69}$$

We obtain Eq. (69) as follows: The gas lying between $z = ut$ and $z = vt$ at time t has density ρ. At time zero this gas occupied the region from $z = 0$ to $z = vt$ with density ρ_o. (The gas beyond the wave front has not been affected.) The gas between the piston and the wave front is conserved. Thus we have

$$\rho(v - u) = \rho_o v \quad . \tag{70}$$

Solving Eq. (70) for u/v gives

$$\frac{u}{v} = \frac{\rho - \rho_o}{\rho} \approx \frac{\rho - \rho_o}{\rho_o} \quad \text{for} \quad u/v \ll 1 \ , \tag{71}$$

which is Eq. (69). (The last equality in Eq. (71) is obtained by writing

$$\rho^{-1} = [\rho_o + (\rho-\rho_o)]^{-1} = \rho_o^{-1}\,[1+(\rho-\rho_o)/\rho_o]^{-1} = \rho_o^{-1}\,[1-(\rho-\rho_o)/\rho_o + \ldots] \approx \rho_o^{-1}.)$$

We are now ready to use Newton's second law to find the velocity v. The rate of change of momentum of the gas at the wave front is given by the product of the amount of mass that goes from velocity zero to velocity u per unit time, $\rho_o Av$, times the acquired velocity u. Thus

$$\frac{d(\text{momentum})}{dt} = (\rho_o Av)\, u \quad . \tag{72}$$

This equals the net force $(p - p_o)\, A$ exerted at the wave front on the fluid that is acquiring momentum. Thus we have (after cancelling the common factor A)

$$(p - p_o) = \rho_o v u \quad . \tag{73}$$

Inserting u from Eq. (69) into Eq. (73) and solving for v^2 we obtain

$$v^2 = \frac{p - p_o}{\rho - \rho_o} = \frac{\gamma p_o}{\rho_o} \ , \tag{74}$$

where the second equality in Eq. (74) is just Eq. (65). For air at STP, p_o

is 1 atmosphere, i.e., 1.01×10^6 dyne/cm^2 ; ρ_o is (29 gm/mole)/(22.4 liter/mole), which is 1.29×10^{-3} gm/cm^3 ; γ is 1.40. Putting these numbers into Eq. (74) we find

$$\begin{aligned} v \text{ (sound in air at STP)} &= 332 \text{ meter/sec} \\ &= 745 \text{ miles/hr} \\ &= 1 \text{ mile}/4.8 \text{ sec.} \end{aligned} \tag{75}$$

The impedance per unit area experienced by the piston when it is emitting a traveling wave is the excess pressure above that necessary to maintain equilibrium per unit velocity imparted to the fluid. We shall call the impedance per unit area by the name Z . The excess force per unit area is the gauge pressure $p - p_o$ exerted by the piston. The velocity imparted to the fluid at the face of the piston is u , the velocity of the piston. Thus according to Eq. (73) the characteristic impedance is given by

$$Z = \frac{p - p_o}{u} = \rho_o v \ , \tag{76}$$

$$Z = (\gamma p_o \rho_o)^{\frac{1}{2}} \ , \tag{77}$$

where we used Eq. (74) to obtain Eq. (77).

Our derivation of the expression for the velocity v and the impedance Z was done by using the simple wave of Fig. 11, but our result is more general than that. Equation (73) can be generalized by replacing p by $p(t)$ and u by $u(t)$, where $p(t)$ is the instantaneous pressure exerted by the piston. Thus Eq. (73) becomes the more general

$$p(t) - p_o = \rho_o v \, u(t) = Z \, u(t) \ , \tag{78}$$

which holds whenever the transmitter is emitting a pure traveling wave.

Let us find the value of Z for air at STP. Using Eq. (76) we have

$$Z(\text{air at STP}) = \rho_o v = 1.29 \times 10^{-3} \times 3.32 \times 10^{4} = 42.8 \text{ gm/cm}^2 \text{ sec.} \quad (79)$$

As a mnemonic, notice that this way of expressing Z shows that Z is the amount of mass per unit area per unit time that is swept through by a traveling wave front.

The flux $I(t)$, defined as the energy per unit area per unit time radiated as traveling waves, is given by

$$I(t) = [p(t)-p_o]\, u(t) = Z\, u^2(t) = [p(t)-p_o]^2/Z \quad (80)$$

A person speaking with an average conversational tone emits about 100 erg/sec of sound energy (and about 100 times that amount when shouting). A commonly used standard of sound flux (energy/cm^2 sec) is

$$\text{standard flux} = I_o = 10 \text{ erg/cm}^2 \text{ sec.} = 1 \text{ micro watt/cm}^2 \ . \quad (81)$$

Thus if we say that the average mouth aperture is about 10 cm^2 when speaking, we see that an average conversational tone gives a flux of about I_o right next to the mouth of the speaker.

Whenever the sound flux increases by a factor of 10 we say that it has increased by 1 bel. A person with normal hearing can barely detect an increase of loudness of about 0.1 bel or 1 decibel, where

$$\begin{aligned} &1 \text{ db} = 1 \text{ decibel} = \text{factor of } 10^{0.1} = \text{factor of } 1.26 \text{ in flux.} \\ &1 \text{ bel} = 10 \text{ db} = \text{factor of } 10 \text{ in flux} \ . \end{aligned} \quad (82)$$

If the flux is 20 to 30 db above I_o (i.e., $I \approx 100$ to 1000 μw/cm^2), the listener begins to feel pain. The flux of the faintest sounds that one can

hear depends on frequency. At about A 440 (i.e., 440 cps) the average person's threshold of audibility is about 100 db less than I_o, i.e., about $10^{-10}\ I_o$. Thus the human ear works over a range of about 10^{12} in intensity!

Here are some examples of numerical calculations of the orders of magnitude of interesting quantities.

1. For sound flux at pain threshold, say at $I = 1000\ I_o$, what is the mean square gauge pressure in atmospheres? We know that if you swim down 10 or 20 feet under the surface of water you feel pain in your ears. One atmosphere is about 33 feet of water, so the water gauge pressure at 15 ft is 1/2 at. (assuming the pressure on the inner side of your eardrum is 1 at.). Is that perhaps the gauge pressure in a sound wave that causes pain?

Solution: We have $[p(t) - p_o]^2 = Z\ I(t)$. [See Eq. (80).] The (given) statement that I is $1000\ I_o$ refers naturally to the time-averaged flux. Thus for the time-averaged value of $(p - p_o)^2$ we have $Z \cdot 1000 \cdot I_o = 42.8 \times 10^4 \approx (650)^2$. Thus the root mean square gauge pressure is 650 dyne/cm^2. One atmosphere is 10^6 dyne/cm^2. Thus we have an RMS (root mean square) gauge pressure that is 6×10^{-4} atmospheres. Clearly the pain is not explained by saying, "It's like swimming under water" -- 6×10^{-4} at. is equivalent to only $\frac{1}{2}$ cm of water!

2. At $I = 1000\ I_o$ (i.e., a painfully loud sound) how does the average fluid velocity (magnitude) compare with the velocity of sound? According to Eq. (71), $u/v = (\rho - \rho_o)/\rho_o \approx \gamma^{-1}\ (p - p_o)/p_o$. Thus the root-mean square value of u/v is roughly equal to that of $(p - p_o)/p_o$, which is 6×10^{-4} at $I = 1000\ I_o$. Thus we have justified our assumption that we have $|u|/v \ll 1$ [Eq. (71)].

3. What is the amplitude of the excursions of the fluid for painfully loud sound? They are the same as the amplitude of the piston. We have

$$D(t) = D_o \cos \omega t$$

$$u(t) = -\omega D_o \sin \omega t \, .$$

$$\langle u^2(T) \rangle = \tfrac{1}{2} \omega^2 D_o^2$$

$$\langle I \rangle = Z \langle u^2 \rangle = \tfrac{1}{2} \omega^2 D_o^2 Z \, .$$

We see that for a given average flux the vibration amplitude is inversely proportional to the frequency, i.e., the amplitude is proportional to the wavelength of the sound. For $I = 100\ \mu$ w/cm^2 (i.e., 1000 erg/cm^2 sec) and $\nu = 440$ cps, we have

$$\text{very loud:} \quad D_o = (2 \cdot 1000/42.8)^{\frac{1}{2}}/(2\pi \cdot 440) = 2.5 \times 10^{-3} \text{ cm.} \quad (83)$$

Note that if we approximate your eardrum by a piston (instead of a drum head), and if there is perfect impedance matching at your eardrum, so that the air "feels" a terminating impedance of 43 gm/cm^2 sec at the eardrum, then the motion of the eardrum is the same as that of the air in a pure traveling wave. In that case, for sound at $100\ I_o$ ($100\ \mu$w/cm^2) your eardrum excursions are equal to D_o as calculated above, i.e., about 2×10^{-3} cm. Of course the response of a drum head is not exactly that of a piston in a tube, and there is also not perfect impedance matching. Nevertheless we expect that Eq. (83) gives the correct order of magnitude for the motion of the eardrum.

4. What is the amplitude of the fluid excursions for barely audible sound? For typical audible frequencies, say 440 cps, one can barely detect a sound flux 100 db below I_o, i.e., $10^{-10}\, I_o$ or $10^{-10}\ \mu\,w/cm^2$. This is 10^{-12} times the flux considered in example 3. The amplitude D_o is proportional to the square root of the average intensity. Thus D_o is

$$\text{barely audible:}\quad D_o \approx 2\times 10^{-9}\ \text{cm} . \tag{84}$$

This is less than the diameter of an average atom, which is $\sim 10^{-8}$ cm. Thus your ear has the fantastic sensitivity of being able to detect motions of the eardrum of the order of less than an atomic diameter! (See also problem 4.)

5. What is the approximate audio (sound) output (in watts) that you would expect to find for a typical radio? Suppose you want to fill a room that has sound absorbing walls with traveling waves carrying flux $100\ I_o$. If the room has cross sectional area $A \approx 10\ \text{ft} \times 10\ \text{ft} \approx 10^5\ \text{cm}^2$, we get for the power P

$$P = 100\, I_o A = 100\ (1\mu\, w/cm^2)\ 10^5\ cm^2 = 10\ \text{watt} .$$

(See also problem 5.)

6. What are the changes of temperature and density in small volume elements of gas in the crests of the wave (or the troughs)? Using Eq. (65), we have

$$\frac{\rho - \rho_o}{\rho_o} = \frac{1}{\gamma}\frac{(p - p_o)}{p_o}$$

From example 1 above, we have that for a painfully loud sound ($I = 1000\ I_o$) the RMS pressure differential is about 6×10^{-4} at. Thus

$$\langle \rho - \rho_o \rangle = (1.29\times 10^{-3})\ (6\times 10^{-4})/(1.4) \approx 6\times 10^{-7}\ \text{gm/cm}^3 .$$

Thus, using $T_o \approx 300^o$ K, we have

$$\langle T - T_o \rangle = (300)\ (.4/1.4)\ (6 \times 10^{-4}) \approx 0.06^o \text{ K}$$

Thus the crests are about $0.06 + 0.06 = 0.1^o$ K hotter than the troughs.

Wave equation and general considerations for sound waves. We have already obtained the wave velocity and the characteristic impedance for sound waves. We also know the dispersion law -- the waves are non-dispersive (for ordinary audio frequencies and intensities). That is all we need. We do not need the wave equation. Nevertheless it is instructive to derive it. In doing so we shall rediscover the expression for the velocity of sound [Eq. (74)] and will find an expression for the flux of sound energy in the $+z$ direction. This expression will be general; i.e., it will include standing waves, traveling waves, or any linear combination of both. As we know, the wave equation also applies to general waves (standing, traveling, or both).

Wave equation. Consider a small amount of fluid that at equilibrium lies between z and $z + \Delta z$. The mass of this fluid is $\rho_o A\, \Delta z$ (A is the cross sectional area of the tube). Its velocity $v(z,t)$ is $\partial\psi(z,t)/\partial t$, and its acceleration is $\partial^2\psi(z,t)/\partial t^2$. Thus

$$\left[\begin{array}{c}\text{mass times acceleration} \\ \text{in } +z \text{ direction}\end{array}\right] = A\Delta z\, \rho_o\, \frac{\partial^2\psi(z,t)}{\partial t^2} \quad . \tag{85}$$

The force in the positive z direction on the left face of the element of fluid is the area A times $p(z,t)$. The force in the positive z direction on the right face is $-Ap(z + \Delta z,t)$. The net force is thus

$$F_z \text{ (on element of fluid)} = -A\,[p(z+\Delta z,t) - p(z,t)] \tag{86}$$

$$= -A\Delta z\,\frac{\partial p(z,t)}{\partial z} - \tfrac{1}{2}A(\Delta z)^2\,\frac{\partial^2 p(z,t)}{\partial z^2} + \cdots \tag{87}$$

$$\approx -A\Delta z\,\frac{\partial p(z,t)}{\partial z} \tag{88}$$

$$= -A\Delta z\,\frac{\partial}{\partial z}\left[p_o - \gamma p_o\,\frac{\partial \psi(z,t)}{\partial z}\right] \tag{89}$$

$$= A\Delta z\gamma p_o\,\frac{\partial^2 \psi(z,t)}{\partial z^2}\quad , \tag{90}$$

where Eq. (87) is the Taylor's series expansion of Eq. (86), Eq. (88) drops the higher derivatives, and Eq. (89) uses the result for $p(z,t)$, Eq. (66). By Newton's second law we can equate the right-hand sides of Eqs. (85) and (90), obtaining the classical wave equation,

$$\frac{\partial^2 \psi(z,t)}{\partial t^2} = v^2\,\frac{\partial^2 \psi(z,t)}{\partial z^2}\quad ,\quad v^2 = \frac{\gamma p_o}{\rho_o}\quad . \tag{91}$$

Energy flux. The wave energy flux S_z in the $+z$ direction (in units of energy per unit area per unit time) is obtained by finding the rate at which work is done on the medium to the right of a point z by the moving parts at z (which may be either fluid or the output of a transmitter). We want to be careful to give only the work that is going into wave motion. That means we want to use the gauge pressure to obtain the force. Thus if the pressure $p(z,t)$ is thought of as a superposition of the equilibrium pressure p_o plus the gauge pressure $p(z,t) - p_o$, it is clear that p_o plays no role in transferring wave energy from one fluid element to the next, because p_o is always pushing on both sides of every little bit of fluid, and never

contributes any net force. The rate of flow of wave energy in the $+z$ direction at point z at time t is thus the force $A[p(z,t) - p_o]$ times the velocity $v(z,t)$ of the fluid. Dividing by the area A, we get the z component S_z of the energy flux vector $\underset{\sim}{S}$ (this is its only component), given by

$$\text{flux vector} = S_z = [p(z,t) - p_o]v(z,t) \tag{92}$$

$$= -\gamma p_o \frac{\partial\psi(z,t)}{\partial z} \frac{\partial\psi(z,t)}{\partial t} , \tag{93}$$

where we used Eq. (7) for $p(z,t)$.

Equation (93) had better be consistent with the expression we obtained in Eq. (78) for the energy flux emitted by a transmitter emitting a pure traveling wave. If you insert into Eq. (93) the proper relation between the space and time derivatives for a pure traveling wave, and consider the boundary conditions at the transmitter, you will see that they are consistent. (See problem 2.)

Pressure and velocity waves. We have derived the wave equation only for the displacement $\psi(z,t)$. The more physically interesting quantities are $p(z,t)-p_o$ and $v(z,t)$, i.e., aside from proportionality constants, $\partial\psi(z,t)/\partial z$ and $\partial\psi(z,t)/\partial t$. It seems intuitively obvious that $p(z,t)$ and $v(z,t)$ must satisfy the classical wave equation as well as $\psi(z,t)$. You may easily show that the wave equation

$$\frac{\partial^2\varphi(z,t)}{\partial t^2} = v^2 \frac{\partial^2\varphi(z,t)}{\partial z^2} \tag{94}$$

is indeed satisfied when $\varphi(z,t)$ is $\partial\psi(z,t)/\partial t$ and when $\varphi(z,t)$ is $\partial\psi(z,t)/\partial z$, provided it is satisfied when $\varphi(z,t)$ is $\psi(z,t)$ (problem 7). In fact, if we know $\partial\psi(z,t)/\partial z$ and $\partial\psi(z,t)/\partial t$ for all z and t

and know $\psi(z,t)$ at a single point, say $z = 0$, for all time [i.e., know $\psi(0,t)$], then we can find $\psi(z,t)$ for all places z and times t (problem 8). We do know $\psi(0,t)$ if we know the motion of the piston. Thus we need not worry about $\psi(z,t)$; it is "known implicitly" if we know the more interesting quantities $p(z,t)$ and $v(z,t)$.

First order coupled wave equations. We have seen that $p(z,t)$ and $v(z,t)$ each obey the second order partial differential wave equation Eq. (94). But that leaves a lot unsaid, because it does not tell us how $p(z,t)$ and $v(z,t)$ are related. A more informative way of expressing the wave nature of $p(z,t)$ and $v(z,t)$ is through the following two coupled first order linear partial differential equations:

$$\rho_o \frac{\partial v(z,t)}{\partial t} = - \frac{\partial p(z,t)}{\partial z} \tag{95}$$

$$\frac{\partial p(z,t)}{\partial t} = - \gamma p_o \frac{\partial v(z,t)}{\partial z} \tag{96}$$

[Equation (95) is equivalent to Newton's law, i.e., to Eqs. (85) through (91). Equation (96) is obtained by first taking the time derivative of the relation [Eq. (66)]

$$p(z,t) - p_o = - \gamma p_o \frac{\partial \psi(z,t)}{\partial z} ,$$

then interchanging the order of the space and time derivatives on the right side, and finally substituting $v(z,t)$ for $\partial\psi(z,t)/\partial t$]. Equations (95) and (96) are intuitively appealing. Equation (95) says that the velocity changes because there is a pressure gradient. If the gradient $\partial p/\partial z$ is positive it means the pressure is greater at greater z, and this will act

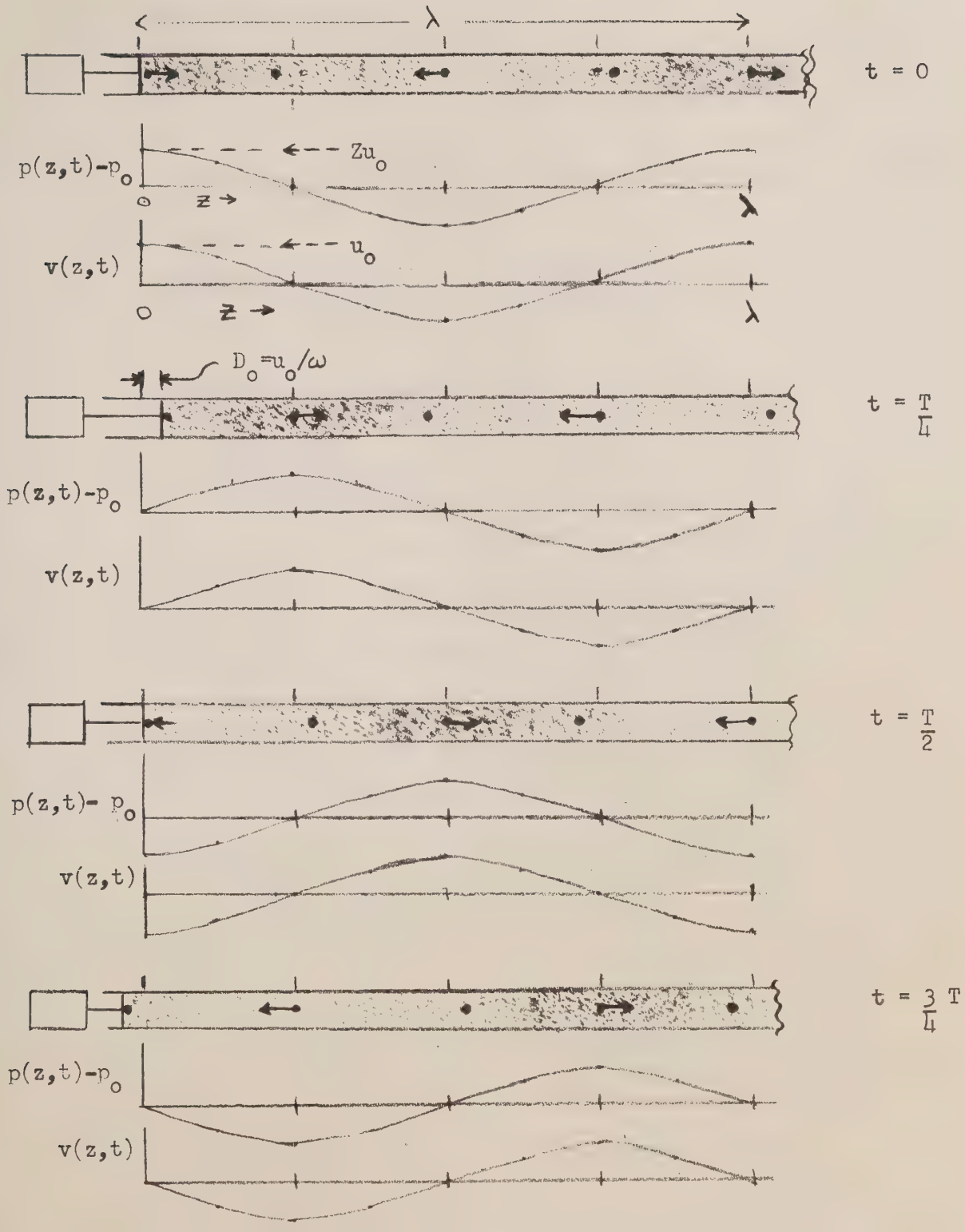


Fig. 3-12. Relations between physical quantities in harmonic traveling sound wave. The small dots follow the motion of single molecules. The amplitude of the piston is taken to be $\lambda/32$.

so as to slow down positive $v(z,t)$; that is why there is a minus sign. The density ρ_o enters in an obvious way -- more inertia means a smaller rate of decrease of $v(z,t)$. Equation (96) says that the pressure changes because there is a velocity gradient. That is also intuitively obvious. The minus sign is there because if the velocity gradient $\partial v/\partial z$ is positive it means fluid with greater z is going away from fluid at z , so the density will decrease at z . Hence the pressure at z will decrease. The pressure decreases at a more rapid rate than does the density by the factor γ , because as the density decreases the temperature drops.

Equations (95) and (96) are the most compact and illuminating expressions for the physics of sound waves. Everything of interest can be derived by starting with these equations as if they came from heaven.

Summary of harmonic traveling sound waves. In Fig. 12 we show the relations that hold in a harmonic traveling wave driven at $z = 0$ by a piston whose velocity and displacement are given by

$$u(t) = u_o \cos \omega t \ , \quad D(t) = (u_o/\omega) \sin \omega t \ . \tag{97}$$

These produce waves satisfying

$$v(z,t) = u_o \cos (\omega t - kz) \ , \tag{98}$$

$$p(z,t) - p_o = Z u_o \cos (\omega t - kz) \ , \tag{99}$$

$$\psi(z,t) = (u_o/\omega) \sin (\omega t - kz) \ . \tag{100}$$

Figure 12 is to be studied very carefully. It contains practically all there is to know about traveling sound waves.

Summary of standing sound waves. In Fig. 13 we show the relations that hold for a harmonic standing wave in an open-ended tube. The quantities plotted

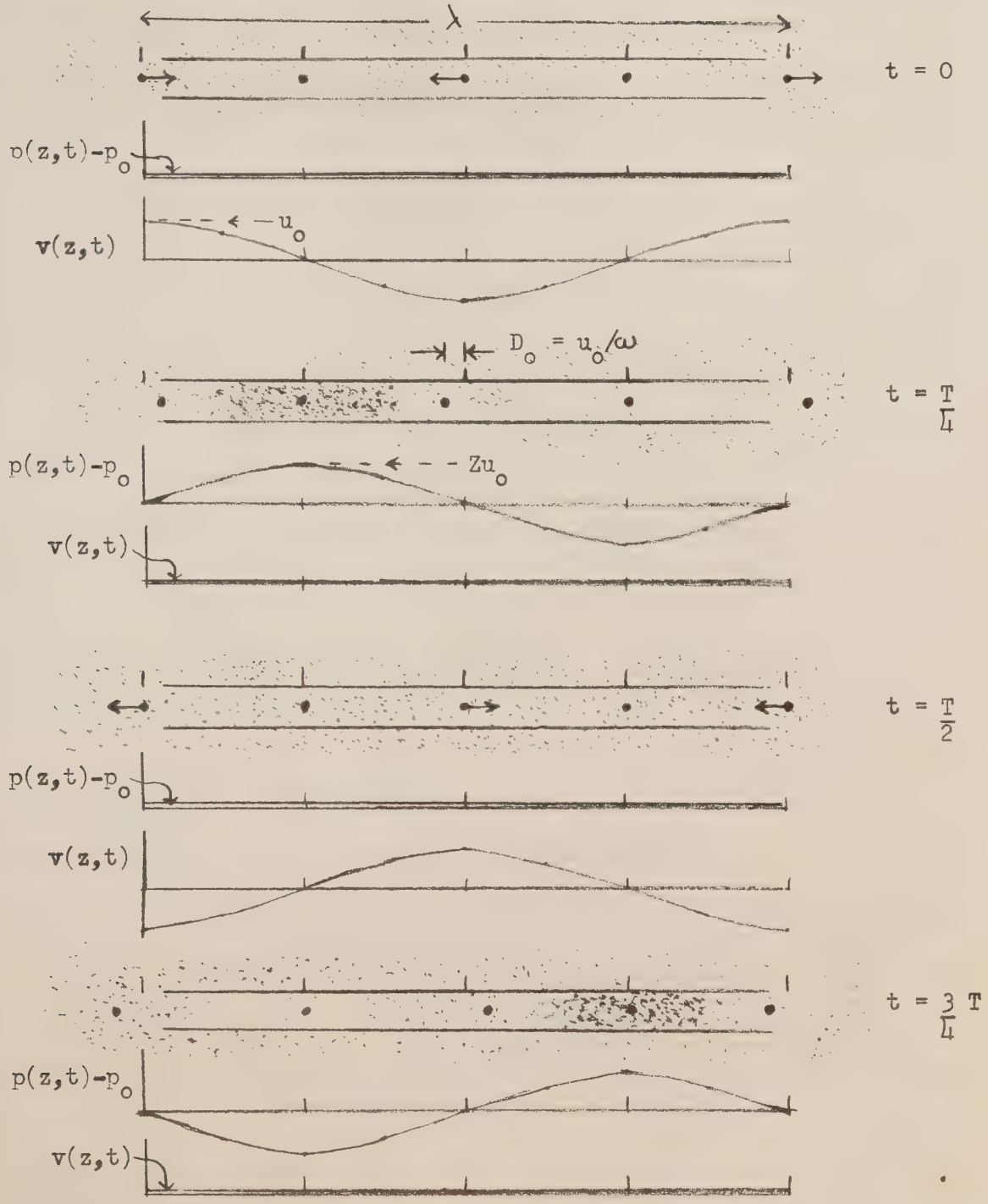


Fig. 3-13. Standing harmonic sound waves in open-ended tube. The small dots follow the motion of individual molecules. The impedance falls to zero outside the ends of the tube within about 0.6 tube radius.

are given by

$$v(z,t) = u_o \cos kz \cos \omega t \ , \tag{101}$$

$$p(z,t) - p_o = Zu_o \sin kz \sin \omega t \ , \tag{102}$$

$$\psi(z,t) = (u_o/\omega) \cos kz \sin \omega t \ . \tag{103}$$

The air at the end of the tube experiences almost no impedance from the room. This can be seen as follows: Think of the last bit of air in the tube as a sort of piston. When this fluid "piston" moves through a distance D_o, the air on the room side of the (imaginary) piston does not have to move a distance D_o, as it would if the tube had the air trapped inside it. Instead it just flows out sideways and "gets away." The impedance does not fall immediately to zero, however. It takes about 0.6 tube radius before the sideways flow of air reduces the gauge pressure to zero. This is shown schematically in Fig. 13, where the tube length is shown slightly shorter than one wavelength for the standing wave (whose wavelength is equal to the "effective" tube length).

Notice that if a barrier were inserted to close the tube at a distance $\lambda/4$ from either (effective) end, the motion of the air would be unaffected. The air molecules at $\frac{1}{4}\lambda$ from the ends are at a node in $\psi(z,t)$; they never move. After inserting this barrier, we could cut off the quarter wavelength end of the tube (including the barrier). Thus we see that the lowest mode of a tube closed at one end has (effective) length of one quarter wavelength.

Similarly, if we cut the tube in two (and then cut off about another 0.6 radius from one end), we would have an open-ended tube in its lowest mode, with an effective tube length of $\frac{1}{2}\lambda$.

Study Figures 12 and 13 carefully. They illustrate Eqs. (95) and (96) and contain nearly everything we wish to tell you about sound.

*3.4 Current and voltage waves on a transmission line.

We wish to find out what happens if at time t = 0 a potential difference V is suddenly applied at z = 0 between the top and bottom edges of a "transmission line" consisting of two parallel conducting planes of width w in the y direction, separated by a gap g in the x direction (with $g \ll w$, so that we can neglect edge effects), and extending forever (or until terminated) in the z direction. The transmission line is shown in Fig. 3-14.

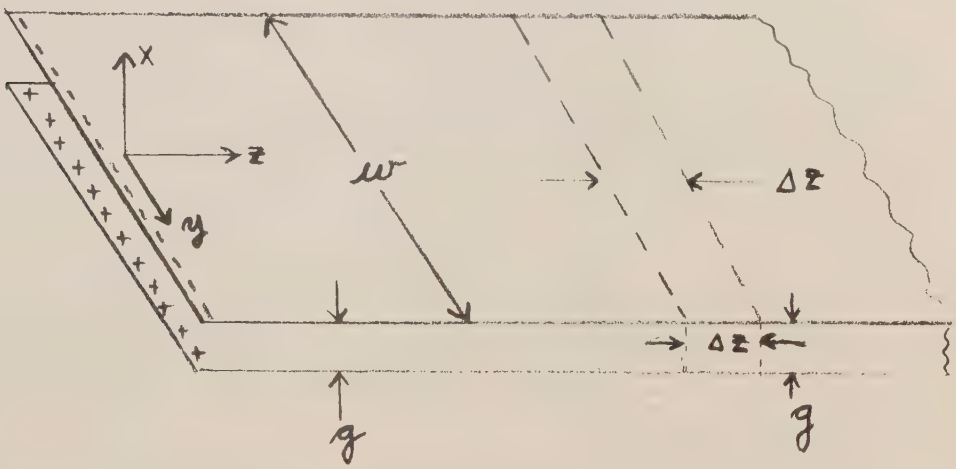


Fig. 3-14. Parallel plate transmission line shown at t = 0, just after a potential difference V has been applied between the top and bottom plates all along the edges at z = 0.

At t = 0 the power supply (not shown) is suddenly connected across the transmission line and thereafter maintains potential difference V. Charge flows in each plate at a velocity v, which we want to calculate, and which we shall assume is constant, independent of frequency. (Our pulse can be considered as a superposition of harmonic waves, as discussed in Chapter 2. We postulate that the velocity of propagation of harmonic waves is the same for each frequency, so there is no spreading of the pulse.

This is the same as assuming that the load is purely resistive, with no reactive component. If our answer for v is in terms only of quantities which are frequency-independent, we will have verified our assumption.) We also want to calculate the load impedance Z felt by the power supply. We can think about the wave as consisting of a step function in charge and current on the conductors, or we can think about the wave as consisting of a step function in the electric and magnetic fields between the plates. We shall do both. (See Fig. 3-15.)

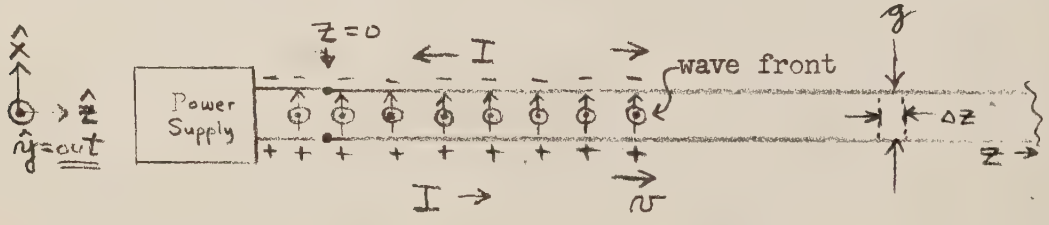


Fig. 3-15. Step function wave propagating at velocity v in parallel-plate transmission line. The electric field is along $\hat{x}$, the magnetic field is along $\hat{y}$. The wave propagates along $\hat{z}$.

From Fig. 3-15 we see that the bottom plate is positive and the top negative. Therefore the electric field between the plates is in the +x direction. We also see that the current in the bottom plate is in the +z direction and that the current in the top plate is in the -z direction. These each contribute to a magnetic field in the +y direction. The electric and magnetic fields are both zero except in the region between the plates (since we are neglecting edge effects). In the region between the plates the fields are perpendicular to one another, and the direction of propagation of the wave front is in the direction of $\underset{\sim}{E} \times \underset{\sim}{B}$, i.e., it is along $\hat{z} = \hat{x} \times \hat{y}$.

Now let us consider in detail what happens at the advancing wave front.

Either we may consider what happens at the edge $z = 0$ between $t = 0$ and the small time interval Δt, or we may consider the advancing wave front at an arbitrary time t. Let us do the latter. We consider what happens at the wave front between time t and time $t + \Delta t$. [We take Δt to be a time interval small enough so that we can neglect all but the first order terms in any relevant Taylor's series. We shall never bother to write any higher order terms which would disappear in the limit $\Delta t \to 0$.] We show the interval Δt schematically in Fig. 3-16.

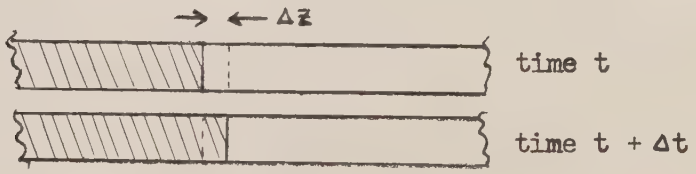


Fig. 3-16. Wave front advancing distance $\Delta z = v\,\Delta t$ in interval Δt.

Now consider the process of "filling up" the region of length Δz with electric and magnetic field during time interval Δt, while the charge density and current build up from zero to their final values. (When we take the limit $\Delta t \to 0$ we shall have "instantaneous" build-up from zero to the final values. Meanwhile we like to look inside the small interval to see what is happening.) The like charges advancing along a plate repel one another. That is why the wave front of charge density advances. The power supply keeps furnishing charge to maintain the process. As the charge wave front reaches our tiny region Δz, a potential difference V is suddenly applied at the front edge of the interval. This potential V tries to build up a current, i.e., it forces the charges into the region Δz. The build-up of

current from zero to I is accompanied by a build-up of magnetic field from zero to the final value of B_y. This process has an "inertia" given by Faraday's law. For a given applied potential V, the rate of current build-up is limited to the value given by

$$L \frac{dI}{dt} = V , \qquad (104)$$

where L is the self-inductance associated with the tiny region Δz in which we are trying to build up the current. Let us define L_z as the self inductance per unit length along z, so that $L_z \Delta z$ is the inductance L of the region Δz. (Later we shall calculate an explicit formula for L_z.)

During the time interval Δt the current in Δz increases from zero to its final value I. The average value of dI/dt is therefore $I/\Delta t$ (neglecting higher order terms, as promised above). The region Δz has length $v \Delta t$, where v is the velocity of propagation of the wave front. Thus Eq. (104) says that $L_z v \Delta t \, I/\Delta t$ is V, i.e.,

$$v L_z I = V . \qquad (105)$$

The larger L_z is, the smaller v is, for a given V and I. The build-up of magnetic field is slower for larger L_z.

After the wave front has passed over Δz, the region Δz has charge +Q on the bottom plate and -Q on the top plate, where Q equals $I \Delta t$. (After the wave front has passed, this is the amount of charge which leaves the right hand edge of Δz in time Δt and which enters the left hand edge.) The capacitance of the region Δz is defined in terms of the potential difference across the gap and the charge displacement Q by the relation

$$V = \frac{Q}{C} . \qquad (106)$$

We define C_z as the capacitance per unit length along z, so that $C_z \Delta z$ is the capacitance C of region Δz. Then in Eq. (106) Q is I Δt, C is $C_z \Delta z$, and Δz is v Δt, so that we have

$$V = \frac{I}{v\, C_z} \quad . \qquad (107)$$

Equations (105) and (107) are each of the form V equals a constant times I Thus each of them says that V and I are related as they are in Ohm's law, so that the impedance offered to V is purely resistive. Thus we have from Eqs. (105) and (107)

$$V = Z\, I \ , \qquad (108)$$

with

$$Z = v\, L_z = \frac{1}{v\, C_z} \quad . \qquad (109)$$

We can solve the second equality of Eq. (109) for v in terms of C_z and L_z. Then we can insert this expression for v into either of the equalities of Eq. (109) to find Z in terms of C_z and L_z. Thus we find

$$v = (C_z{}^{-1}/L_z)^{\frac{1}{2}} \qquad (110)$$

$$Z = (C_z{}^{-1}\, L_z)^{\frac{1}{2}} \quad . \qquad (111)$$

Note that we have verified our assumption of constant velocity of propagation.

We shall shortly calculate C_z and L_z explicitly for the parallel-plate transmission line. Meanwhile we wish to point out that we have not made explicit use of the geometry, except to demand that it be independent of z. Thus Eqs. (110) and (111) hold, for example, for a coaxial transmission line as shown in Fig. 17.

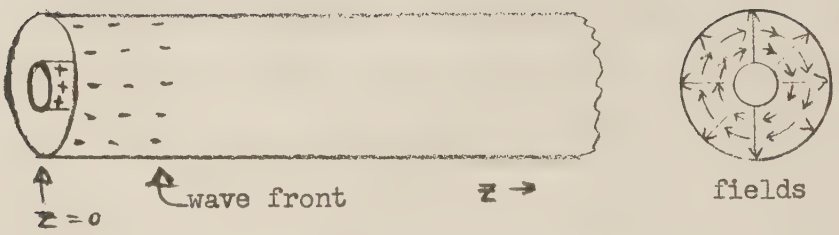


Fig. 3-17. Coaxial transmission line. Current I flows in the +z direction on the inner conductor and in the -z direction on the outer conductor. The electric field is radial and directed outwards. The magnetic field goes in circles that are clockwise (looking in the +z direction). The magnetic field is everywhere perpendicular to the electric field, and $\underset{\sim}{E} \times \underset{\sim}{B}$ is along the propagation direction $\hat{z}$.

If the medium between the two conductors is not vacuum but is a linear homogeneous isotropic medium with dielectric constant ϵ and magnetic permeability μ , then the capacitance C is larger than it would be for vacuum by a factor ϵ , and the inductance L is larger by a factor μ . Thus if we use a subscript zero to denote quantities that apply when there is vacuum between the conductors, we can write Eqs. (110) and (111) in the form

$$v = (\epsilon^{-1}/\mu)^{\frac{1}{2}} v_o \quad , \tag{112}$$

$$Z = (\epsilon^{-1} \mu)^{\frac{1}{2}} Z_o \quad , \tag{113}$$

where v_o and Z_o are the propagation velocity and impedance for vacuum. You may have noticed that Eq. (112) is equivalent to the very general Eq. (132) of Chap. 2, provided that v_o in Eq. (112) is equal to c . That must be the case because our treatment in Chap. 2 was so general that it includes our present example as a special case. Indeed we shall show shortly that

[using Eq. (110)]

$$v_o = (C_z^{-1}/L_z)_o^{\frac{1}{2}} = c \ , \qquad (114)$$

both for the parallel-plate transmission line and for the coaxial transmission line. (We shall not prove the general case. That is easily done, and of course it is included in our general result in Chap. 2.)

The characteristic impedance Z_o does depend on the geometry. It differs between the parallel-plate and the coaxial transmission lines. To put it differently, the amount of current I that flows when a potential V is applied (across the two conductors of a transmission line) depends on the cross sectional geometry (for example, on whether we have a parallel-plate or a coaxial line). The propagation velocity of the wave does not. To put it still another way: the electric field in the wave is proportional to the potential difference V that propagates down the line; V is in turn proportional to C^{-1}, according to Eq. (106). The magnetic field is proportional to L. Thus the fact that C^{-1}/L is independent of the geometry [i.e., Eq. (114) holds] means that, for vacuum, the ratio of electric to magnetic fields is independent of the geometry. The product of the fields is proportional to $C^{-1}L$, and that does depend on the geometry. We shall look at the fields in more detail shortly. Then we shall find that the electric and magnetic fields are always perpendicular to each other and to the direction of wave propagation and that in a pure traveling wave the electric and magnetic fields are at all times and places equal in magnitude (this holds in electrostatic units -- more generally, the electric and magnetic energy densities are equal in any units) and their directions are always related in such a way that the propagation direction is along $\hat{E} \times \hat{B}$.

(These statements refer to pure traveling waves only.)

Reflection at resistive termination. If the transmission line is terminated by a resistor having resistance equal to the characteristic impedance Z given by Eqs. (111) and (113), the wave will be terminated without reflection. The power supply will continue to emit current I at voltage V and will experience a purely resistive load of impedance Z. If the terminating resistor has resistance Z_2, and if the characteristic impedance of the line is Z_1, there will be a reflected wave if Z_2 does not equal Z_1. The reflection coefficient was derived in Sec. 3.2, and is given by

$$R_{12} = \frac{Z_1 - Z_2}{Z_1 + Z_2} \quad . \tag{115}$$

We must now exercise some care. Remember that for waves on a string and for sound waves the reflection coefficient for the waves of the velocity $v(z,t)$ of the displaced particles of the medium is given by

$$R_{\partial\psi/\partial t} = R_{12} \quad , \tag{116a}$$

whereas the reflection coefficient for the transverse tension waves (in the case of the string) or the pressure waves (for sound) is given by

$$R_{\partial\psi/\partial z} = - R_{12} \quad . \tag{116b}$$

In the present example of the transmission line we have two physical quantities of interest, namely V and I, or if you prefer, the corresponding fields E_x and B_y. It should not surprise us to find that one of them reflects with coefficient R_{12} and the other with coefficient $-R_{12}$. In order to see which does which, let us consider an extreme case where it is

obvious what happens. Let us consider the case where Z_2 is zero -- a "short circuited" termination at which the two conductors are connected by a slab of copper (for example) having negligible resistance. Then according to Eq. (115) we have $R_{12} = +1$. All we have to do is use our physical knowledge to see what the current and potential do and thus to see which of them goes with $+R_{12}$ and which with $-R_{12}$. See Fig. 18.

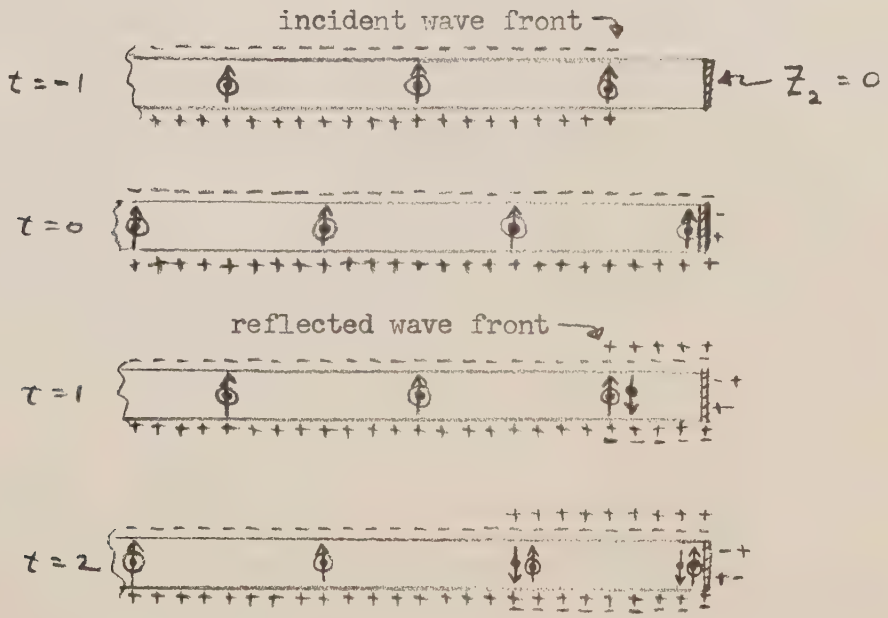


Fig. 3-18. Reflection at resistive load of zero impedance. The potential V and the electric field E_x are reflected with a sign reversal; the current I and magnetic field B_y are reflected without sign reversal. The total potential and electric field are zero to the right of the reflected wave front. The total current and magnetic field to the right of the reflected wave front are twice what they would be for perfect termination.

When Z_2 is zero the charges(of either sign) go "around the corner" and continue back towards the transmitter on the other conductor. They are carried around the corner by their inertia and by the push from the like charges behind. (As far as the charges of one sign are concerned, they are being pushed just as they were before, since charges of the opposite sign are equally distributed before and behind.)

The superposition of incident and reflected waves at the termination gives zero for V and E_x and gives $2I$ for the current and $2B_y$ for the magnetic field. It is easy to remember this without the details of Fig. 18: since we have $V = ZI$, zero impedance means zero potential even for finite current I. The way you get zero for V is to superpose a negative reflected wave of potential upon the incident wave. To summarize:

$$R_V = R_E = -R_{12} = -\frac{(Z_1 - Z_2)}{Z_1 + Z_2} \tag{117a}$$

$$R_I = R_B = +R_{12} = +\frac{(Z_1 - Z_2)}{Z_1 + Z_2} \quad . \tag{117b}$$

By comparison of Eqs. (117) with Eqs. (116), we see that the potential and electric field reflect in the same way as the transverse tension (for waves on a string) or the pressure (for sound waves). Thus we may wish to think of the electric field as a "return-force-like" field. Similarly we see that the current and magnetic field reflect in the same way as the transverse velocity (for waves on a string) or longitudinal velocity (for sound waves). Thus the magnetic field is a "velocity-like" field (it is due to the velocity of the moving charges). Pursuing this analogy (or mnemonic), we notice that in a dispersive medium E is proportional to ϵ^{-1} and B is proportional

to μ (for given sources of current and charge). We see that ϵ^{-1} in Eqs. (112) and (113) for v and Z is analogous to string tension T or gas pressure p_o, whereas μ is analogous to mass density. (We should not try to push these analogies too hard. They are no substitute for an honest derivation.)

*Calculation of L_z and C_z for parallel-plate transmission line. To calculate B_y we use the fact that when there is vacuum between the plates, $\nabla \times \underset{\sim}{B}$ is $4\pi c^{-1}\underset{\sim}{J}$, and hence the line integral of $\underset{\sim}{B}$ around any closed path is $4\pi c^{-1} I$, where I is the enclosed current. For the parallel-plate line of Fig. 14 we take a path of length w along $+y$ between the plates and then go back along $-y$ outside of the plates where $\underset{\sim}{B}$ is zero. This gives $w B_y$ for the line integral, and thus

$$B_y = \frac{4\pi I}{c\, w} \quad . \tag{118}$$

The self-inductance L for a length Δz is given by Faraday's law in the form which states that LdI/dt equals c^{-1} times the rate of change of magnetic flux through the area $g\Delta z$ (see Fig. 14). This rate of change of flux is dB_y/dt times the area. Thus, according to Eq. (118), the rate of change of flux through this area is $4\pi(dI/dt)g\Delta z/cw$, so that LdI/dt is $c^{-2}\, 4\pi(g/w)\, \Delta z\, dI/dt$. By definition, $L/\Delta z$ is L_z. Thus

$$L_z = c^{-2}\, 4\pi\, (g/w) \quad . \tag{119}$$

The electric field is given by Gauss's law to be 4π times the charge density σ, which is the charge Q on the strip of area $w\Delta z$ divided by that area. Thus

$$E_x = 4\pi\sigma \tag{120}$$

$$= 4\pi\, Q/w\Delta z \ . \tag{121}$$

The potential difference V between the plates is $E_x g$. It is also equal to $C^{-1}Q$, which equals $C_z^{-1}\, Q/\Delta z$, by definition of C_z. Thus according to Eq. (121) we have

$$C_z^{-1} = 4\pi(g/w) \ . \tag{122}$$

Inserting Eqs. (119) and (122) into Eqs. (112) and (113) gives (for vacuum)

$$v_o = (C_z^{-1}/L_z)^{\frac{1}{2}} = c \tag{123}$$

$$Z_o = (C_z^{-1}L_z)^{\frac{1}{2}} = \frac{4\pi}{c}\frac{g}{w} \ . \tag{124}$$

Since we are using electrostatic units, the impedance Z_o is given here in statohms, which have dimensions sec/cm. One statohm is 1 statvolt/1 statamp = (300 volts)/[(1/3) $\times$ 10^{-9} amp] = 30 $\times$ 3 $\times$ 10^{10} Ω (ohms). Thus c^{-1} statohm is $(3 \times 10^{10})^{-1}$ statohm or 30 Ω. Thus

$$\frac{4\pi}{c} \text{ statohms} = 120\ \pi\Omega = 377\ \Omega \tag{125}$$

Thus Eq. (124) can be read as

$$Z_o = 377\,\frac{g}{w}\,\Omega \ . \tag{126}$$

For example, a parallel-plate transmission line of gap 2mm and width 2 cm has characteristic impedance 37.7 ohms (neglecting edge effects).

* Calculation of L_z and C_z for coaxial transmission line. We leave it to you to show [see Vol. 2, p 185, Eq. (2)] that for the coaxial transmission line corresponding to Fig. 17 with vacuum between the conductors we have the

c

azimuthal magnetic field B_ϕ given by

$$B_\phi = \frac{2}{c}\frac{I}{r} \quad , \tag{127}$$

between $r_{min} = a$ and $r_{max} = b$, the radii of the inner and outer conductors, with B zero elsewhere. This gives a self-inductance per unit length (as you may show) of

$$L_z = 2\, c^{-2} \ln(b/a) \quad . \tag{128}$$

The electric field is radial and given by [see Vol 2, p. 28, Eq. (27)]

$$E_r = \frac{2(Q/\Delta z)}{r} \quad , \tag{129}$$

where Q is the charge on one of the conductors in the length Δz . The potential difference V is then $2(Q/\Delta z)\ln(b/a)$. This also equals $C_z^{-1}\, Q/\Delta z$, by definition of C_z . Thus

$$C_z^{-1} = 2 \ln(b/a) \quad . \tag{130}$$

The propagation velocity and characteristic impedance are then

$$v_o = (C_z^{-1}/L_z)^{\frac{1}{2}} = c \quad ,$$

$$Z_o = (C_z^{-1} L_z)^{\frac{1}{2}} = \frac{2}{c} \ln(b/a) \quad ,$$

In MKS units, we have for the coaxial cable

$$Z_o = 60 \ln(b/a) \text{ ohms} \quad .$$

Thus a coaxial cable for which the inner conductor has radius 1/2.718 times that of the outer conductor has a characteristic impedance of 60 ohms. We have illustrated the fact that the propagation velocity is always c (for

vacuum) but the impedance depends on the geometry.

The fields. The current I is the charge per unit length, $Q/\Delta z$, times the velocity of propagation v of the wave. When there is a dispersive medium with permeability μ and dielectric constant ϵ, the magnetic field is given by μ times the vacuum expressions Eq. (118) or (127). Also, the electric field is given by ϵ^{-1} times the vacuum expressions Eq. (121) or (129). In either the parallel-plate or the coaxial transmission line the electric and magnetic fields are perpendicular to one another and to the direction of propagation, which is along $\hat{\underset{\sim}{E}} \times \hat{\underset{\sim}{B}}$. The ratio of the magnitudes of $\underset{\sim}{E}$ and $\underset{\sim}{B}$ is also the same in both cases:

$$\frac{|\underset{\sim}{B}|}{|\underset{\sim}{E}|} = \frac{\mu}{\epsilon^{-1}} \frac{v}{c} = (\epsilon\mu)^{\frac{1}{2}} = \frac{c}{v} = n \quad . \tag{131}$$

Thus in vacuum $\underset{\sim}{B}$ and $\underset{\sim}{E}$ have the same magnitude. These relations between $\underset{\sim}{B}$ and $\underset{\sim}{E}$ hold for any pure traveling electromagnetic wave, no matter what the geometry.

Energy flux. Let us consider the energy balance for the parallel-plate line. The power P emitted by the transmitter is the voltage V times the current I. The power per unit cross sectional area of the traveling wave is thus $P/(\text{area})$, i.e. VI/gw. The potential V is the electric field E_x times the gap g, so that V/g is E_x. The ratio I/w is $(c/4\pi)B_y$ in vacuum, according to Eq. (118), and is more generally $(c/4\pi)(B_y/\mu)$. Thus the energy flux in the traveling wave, S_z, is given (in erg/cm^2 sec) by

$$S_z = \frac{c}{4\pi} E_x (B_y/\mu) \quad . \tag{132}$$

Equation (132) is a special case of a more general relation that holds for

any electromagnetic wave in a dispersive medium:

$$\underset{\sim}{S} = \frac{c}{4\pi} \underset{\sim}{E} \times (\underset{\sim}{B}/\mu) \quad . \tag{133}$$

Equation (133) holds for the waves propagating in the coaxial transmission line, as well as for those propagating in the parallel-plate transmission line, and in fact it holds in general, not only for pure traveling waves, but also for standing waves or for any combination of standing or traveling waves. We can see that this is true for the special case of the parallel-plate line, Eq. (132), as follows: Instead of considering V and I at the transmitter, consider them at any point z at time t. The sign conventions on V and I are such that VI gives the rate at which the fields at z are storing energy in the region to the right of z. Thus the expression $VI/(\text{area})$ gives the energy flux in the $+z$ direction for any superposition of fields. By substituting $E_x g$ for V and using Eq. (118) for I in this expression, we obtain Eq. (132). For a more complicated geometry, as in the case of the coaxial line, we can still apply Eq. (133) in a small region (i.e., a region with small dimensions transverse to the propagation direction $\hat{z}$, so that we may regard the fields as constant in the region).

Plane waves in unbounded space. All of our results that can be expressed "locally" in terms of the fields $E_x(z,t)$ and $B_y(z,t)$ are independent of the particular sources of charge and current that generated those fields. To put it differently, one cannot tell by examining the electromagnetic fields in a limited region of space what charge and current distributions produced those fields. By using the superposition principle one can find an infinite number of different charge and current distributions that will produce those same field configurations. In our parallel-plate example

we found a wave of electric and magnetic fields $E_x(z,t)$ and $B_y(z,t)$ propagating in the z direction with E_x and B_y independent of lateral position x and y (as long as we stay between the two conductors). Even that restriction can be removed -- we simply suppose that the plates are very wide and have a very large gap -- as wide and large as we please. The fields "can't tell the difference." (Of course the power supply has to put out more power to fill more space with fields; but the fields at z,t don't know about that.) There are many situations for which traveling plane waves (i.e., traveling waves whose amplitudes and phases do not depend on the transverse dimensions x and y) are a good approximation. For example, a small radio antenna on a satellite in interplanetary space emits electromagnetic waves that propagate out in all directions. (More likely it will have an antenna which emits preferentially towards the earth.) Far from the antenna, in any limited region of space, the waves are essentially plane waves. They then have all the properties we have just described;

Traveling plane waves

$$\underset{\sim}{E} \cdot \underset{\sim}{B} = 0 \quad \text{at every z and t} \ , \tag{134a}$$

$$\hat{\underset{\sim}{E}} \times \hat{\underset{\sim}{B}} = \text{propagation direction} \equiv \hat{z} \ , \tag{134b}$$

$$|\underset{\sim}{B}| = n\,|\underset{\sim}{E}| \quad \text{at every z and t} \ , \tag{134c}$$

$$\underset{\sim}{S} = \frac{c}{4\pi}\,\underset{\sim}{E} \times (\underset{\sim}{B}/\mu) \ , \tag{134d}$$

$$\text{energy density} = \frac{1}{8\pi}\left\{\epsilon \underset{\sim}{E}^2 + \mu^{-1}\,\underset{\sim}{B}^2\right\} \ , \tag{134e}$$

where the last equation, Eq. (134e), is easily obtained for the case of traveling waves from the other equations and from the fact that the energy flux for traveling waves is v times the energy density. (The two contributions $\epsilon\,\underset{\sim}{E}^2$ and $\mu^{-1}\,\underset{\sim}{B}^2$ are equal in a traveling wave at every z and t .

In a pure standing wave $\underset{\sim}{B}^2$ is zero when $\underset{\sim}{E}^2$ is maximum, and vice versa, The maxima of $\underset{\sim}{B}^2$ occur at different places from the maxima of $\underset{\sim}{E}^2$. The behavior of E_x and B_y in both traveling and standing waves is analogous to that of the pressure and the velocity in sound waves.)

* Reflection of light waves at a discontinuity in index of refraction. Now that we know (or at least have claimed) that our results for the parallel-plate transmission line are actually of much greater generality, we shall boldly calculate the reflected intensity for light incident on a surface of discontinuity in the index of refraction, as,for example, for visible light incident from air to glass or glass to air. Let us calculate the reflection coefficient for the electric field. (We know that the magnetic field has the opposite sign and same magnitude for its reflection coefficient, and therefore we need not calculate it separately.) We think of all space as being one great big parallel-plate transmission line. When the dielectric constant ϵ and magnetic permeability μ change at the plane of discontinuity (which is a plane perpendicular to z), the geometry stays the same(i.e., in transmission-line terms, we want to keep the width w and gap g constant). Now, Eq. (117a) gives for the reflection coefficient R_E of the electric field

$$R_E = -R_{12} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{(Z_2/Z_1) - 1}{(Z_2/Z_1) + 1} \quad . \tag{135}$$

Now, aside from quantities that are the same for both media (and hence cancel in the ratio Z_2/Z_1), the impedance Z of a dispersive medium is given by [see Eq. (113)]

$$Z = \text{const} \cdot (\mu/\epsilon)^{\frac{1}{2}} = \text{const} \cdot \mu \frac{v}{c} = \text{const} \cdot \mu v \quad . \tag{136}$$

Since we wish to consider a dispersive medium, we may have ϵ and μ dependent on the frequency ω. Therefore we write ω/k in place of v, and then absorb the frequency ω into the constant. We can do this because the frequency does not change when we go from one medium to another. (It just wouldn't make any sense at all for the frequency to change! The whole system, including all the different media, is just one big driven oscillator, driven at frequency ω. Thus, no matter what its natural "resonance" frequencies (i.e., normal modes) might be, everything settles down into a steady state at the driving frequency ω.) Of course the wavelength and wave number do change in going from one medium to another. Thus we have

$$Z = \text{const} \cdot \frac{\mu(k)}{k} \quad . \tag{137}$$

It turns out to be more convenient to write

$$Z^{-1} = \text{const} \cdot (k/\mu) \quad , \tag{138}$$

or

$$Z^{-1} = \text{const} \cdot (n/\mu) \quad . \tag{139}$$

Thus Eq. (135) can be put in the forms

$$R_E = \frac{(n_1/n_2) - (\mu_1/\mu_2)}{(n_1/n_2) + (\mu_1/\mu_2)} = \frac{(k_1/k_2) - (\mu_1/\mu_2)}{(k_1/k_2) + (\mu_1/\mu_2)} \quad . \tag{140}$$

Here is an important practical application of Eq. (140). For light incident from air to glass, we may take as sufficiently good approximations $n_1 = 1$ (air) and $n_2 = 1.5$ (glass). The magnetic permeability may be taken to be 1 for both air and glass, to a very good approximation. Then Eq. (140) gives

$$\begin{aligned} R_E &= -0.20 \text{ (air to glass)} \\ &= +0.20 \text{ (glass to air)} \quad . \end{aligned} \tag{141}$$

The reflected energy flux (usually called the intensity) averaged over one period of oscillation is R_E^2 times the incident flux, i.e., 0.04 times the incident flux. Thus after going through the two air-glass and glass-air surfaces of a lens, the light emerges from the exit surface with an intensity loss of about 8% (about 4% from each surface). This loss can be very troublesome in optical instruments. (Often it is not the loss but the glare from the reflected light that is bothersome.) It can be largely eliminated by a method called optical coating. Thus advertisements for good lenses and cameras always say "fully coated optics," or something equivalent. See problem 17 for the explanation of how coating works.

*3.5 Radiation by an oscillating sheet of charge.

We shall give one more example of plane electromagnetic waves of the form $E_x(z,t)$ and $B_y(z,t)$ propagating in the $+z$ direction. The fields satisfy the general relations given by Eqs. (134), as they must. However the particular radiating charge and current distribution that we now consider is one that is much closer to the situations encountered in radio broadcasting or in emission of light by oscillating charges. In our example of the parallel-plate transmission line, we thought about letting the plates be very far away ("at infinity") and concluded that there would be no detectable difference between the resulting plane waves and those emitted by a distant light bulb. But there is certainly a great difference when we include the sources, i.e., there is a great difference between a distant light bulb and great sheets of charge and current racing along "at infinity."

We imagine a sheet of charge with uniform positive charge density σ (in electrostatic units of statcoul/cm^2) occupying the entire x-y plane at $z = 0$. The sheet is so thin that we can neglect any effects of its thickness.

First consider the situation at equilibrium, with all the charges at rest. For positive z there is a uniform electric field in the $+z$ direction with magnitude equal to $2\pi\sigma$, according to Gauss's law. [Vol. 2, p. 29, Eq. (28)] For negative z there is a field of equal magnitude and opposite direction, i.e., in the $-z$ direction. The magnetic field is zero everywhere at equilibrium. This situation is shown in Fig. 19.

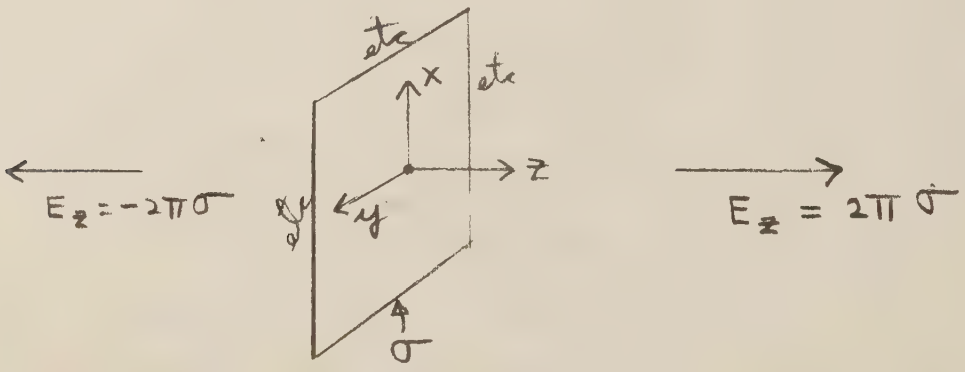


Fig. 3-19. Infinite sheet of charge of zero thickness, covering the entire x-y plane at $z = 0$ with a uniform charge density σ.

Now let us suppose that some force drives the charges along the plus or minus x direction so that they all undergo the same displacement $D_x(t)$ from their equilibrium positions. Let the corresponding velocity of the charges in the sheet be $u_x(t)$. Thus

$$D_x(t) = \text{displacement of charge from equilibrium} \tag{142}$$

$$u_x(t) = \text{velocity of charge} = \frac{d\,D_x(t)}{dt} \tag{143}$$

Since each of the charges undergoes the same motion, just as much charge will flow into a given region (on the sheet at $z = 0$) as will flow out. Thus the charge density σ remains constant. Therefore the component of electric field that is perpendicular to the sheet, namely E_z, remains constant at its equilibrium value of $2\pi\sigma$. This follows directly from Gauss's law, which holds whether or not the charges are in motion. Thus for all z and t we have the constant fields

$$E_z(z,t) = 2\pi\sigma \quad , \quad z > 0 \quad , \quad \text{all} \quad t \tag{144}$$

$$E_z(z,t) = -2\pi\sigma \quad , \quad z < 0 \quad , \quad \text{all} \quad t \quad . \tag{145}$$

The electric field component along $\hat{x}$ is not constant. We can think of a line of force as being attached to a given charge. Then an oscillation of the charge along the plus or minus x direction will induce "kinks" in the lines of force, which we expect to propagate outwards from the sheet at the velocity of light. We shall assume that that is what happens. (We hope that familiarity is breeding complacency about the fact that when you shake something which is coupled to a medium that can support waves, you get waves.) There will be a wave of magnetic field as well as of electric field, because the motion of the charges constitutes a current, which induces a magnetic field. Since it is the same charges which create both the electric and the magnetic fields, we expect a definite relation between the electric and magnetic fields, and of course we shall find one.

It would be nice if we could use the same kind of simple step function in the velocity of the charges in the sheet that we used for generating waves on the string, in the tube of air, and on the transmission line. Unfortunately that will not work in the present example for a very interesting reason. We

want to be able to use the "one-dimensional" aspect of the geometry. That is, we want to use Gauss's law and the symmetry of the infinite sheet to find a simple answer. But if we start all the charges moving together at $t = 0$, we get an answer that is not simple. Instead it is terribly complicated because of the different times for propagation of field changes from different parts of the sheet to the place where we want to measure E_x. The contributions from different parts of the sheet do not all arrive at once, and we have a transient of complicated nature at the field point. You might say, "Wait until the transients have swept past, i.e., until the distant parts of the sheet have all communicated the fact that they are moving at constant velocity." But in that case, the field E_x is back to zero again! That is because a charge sheet moving at steady velocity has a field exactly like the electrostatic one, namely a field $E_z = 2\pi\sigma$ perpendicular to the surface. [See Vol. 2, p. 157, Fig. 5.9(d), which shows a hypothetical but wrong field (as the author shows) for a moving sheet of charge. The lines of force do not "drag behind."] We did not get into this kind of trouble with the transmission line, because the charges and currents propagated out along with the wave front and maintained the "squareness" of the geometrical relations.

We want a situation that enables us to be in steady-state equilibrium with the entire sheet, so that we can make use of the symmetry. There is a way to accomplish this (aside from the uninteresting cases of zero velocity or constant velocity, which do not give any waves). That way is to have the whole sheet undergo a pure harmonic oscillation that goes on and on forever. There is plenty of time for all parts of the sheet to get into equilibrium with the field point. Then we can use the symmetry and say that Gauss's

law is equivalent attaching the lines of force to the charges, and we need only worry about kinks propagating along the z axis. By using a steady state harmonic oscillation we will eliminate the worry about kinks propagating transverse to $\hat{z}$.

After these preliminary remarks, the rest is easy. In Fig. 20 we show one of the oscillating charges of the sheet with its line of force attached to it. From the picture it is obvious that the x component of $\underset{\sim}{E}$ is

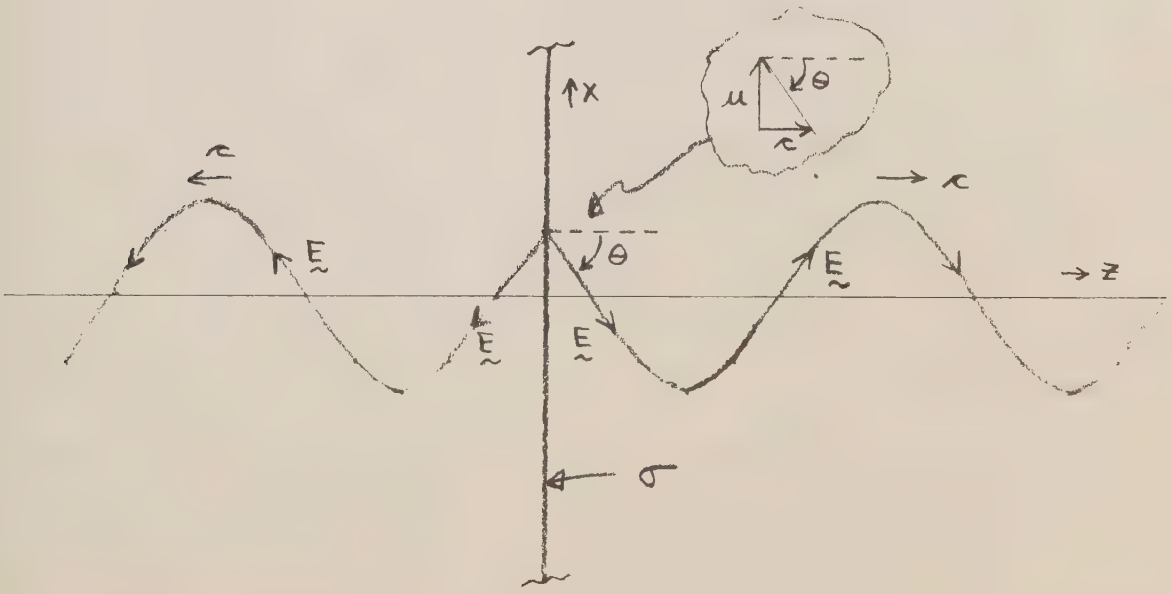


Fig. 3-20. Oscillating charge sheet induces outward-propagating bends in lines of force.

negative (at the instant shown) and is given by

$$E_x = - E_z \tan \theta \quad . \tag{146}$$

But we also see from the figure that $\tan \theta$ is given by

$$\tan \theta = u_x/c \quad . \tag{147}$$

Thus by combining Eqs. (146) and (147), with $E_z = 2\pi\sigma$ [Eq. (144)], we obtain

$$E_x(0,t) = -2\pi\sigma\, u_x(t)/c \quad , \tag{148}$$

which implies

$$E_x(z,t) = -2\pi\sigma\, u_x(t')/c \quad , \tag{149}$$

where t' is the "retarded" time

$$t' = t - \frac{z}{c} \quad . \tag{150}$$

The magnetic field is obtained by considering the current in the sheet. Consider a swath of width w in the y direction. The total current I_x in this swath is the charge density σ times the width w times the velocity $u_x(t)$ of the charges. The current I_x is therefore given by

$$I_x = w\sigma\, u_x(t) \tag{151}$$

as shown in Fig. 21.

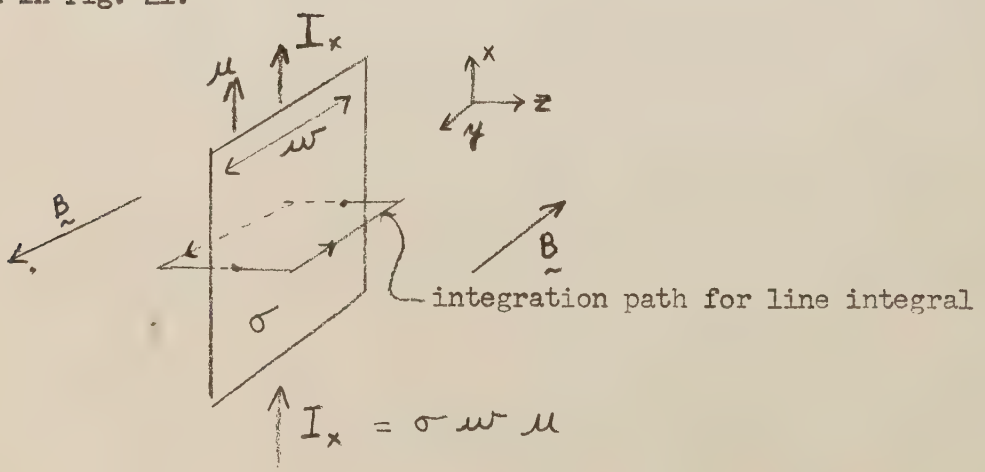


Fig. 3-21. Geometrical considerations for finding the magnetic field.

The line integral of $\underset{\sim}{B}$ around the path shown in Fig. 21 is $-2wB_y$, where B_y is the (negative) value of the y component of $\underset{\sim}{B}$ for positive z

just outside the current sheet. This line integral equals $4\pi/c$ times the current I_x. Thus $-2wB_y$ is $4\pi/c$ times $w\sigma u_x$, and we find

$$B_y(0,t) = -2\pi\sigma u_x(t)/c \quad , \tag{152}$$

which implies

$$B_y(z,t) = -2\pi\sigma u_x(t')/c \quad , \tag{153}$$

where t' is the retarded time.

We see that $B_y(z,t)$ and $E_x(z,t)$ are equal, as they must be for a plane traveling wave propagating in the $+z$ direction.

Energy considerations. Let us calculate the work done by the power supply that pushes the charges in our sheet-like antenna. Consider an area A of dimensions Δx along x and w along y. The field exerted by the power supply on the charges must overcome the force exerted on the charges by the radiated field just outside the sheet. Thus the potential difference through which the power supply drives the charges in a distance Δx is $-E_x \Delta x$, where E_x is the radiated field at $z = 0$:

$$V_x = -E_x \Delta x \quad . \tag{154}$$

The current I_x is $w\sigma u_x$ [Eq. (151).] Therefore the power spent by the supply is

$$P = V_x I_x = -E_x(\sigma u_x)(w \Delta x) \tag{155}$$

$$= -E_x\left(\frac{-c}{2\pi} E_x\right) A = \frac{c E_x^2}{2\pi} A \quad . \tag{156}$$

Half of the power expended by the power supply is radiated in the $+z$ direction and half is radiated in the $-z$ direction. Thus the energy flux radiated in the $+z$ direction, evaluated at $z = 0$ and time t, is

$$\text{flux} = \frac{P}{A} = \frac{c\,E_x^2(0,t)}{4\pi} \quad . \tag{157}$$

That means that a general z,t we have the flux

$$S_z(z,t) = \frac{c\,E_x^2(z,t)}{4\pi} \quad . \tag{158}$$

Our results agree with the general case given by Eqs. (134), as they must.

What happens if the charges in our sheet oscillate along $\hat{z}$ instead of along $\hat{x}$ or $\hat{y}$? There is no radiation. Gauss's law assures us that there are no surges of variation in the z component of electric field. Thus the lines of force do not act like stretched strings along which we may send longitudinal waves of variations of tension.

3.6 The general case of electromagnetic plane waves in vacuum.

This section will be rather abstract. Its purpose is to show that the results we obtained from our examples of plane electromagnetic waves in parallel-plate and coaxial transmission lines and from a radiating sheet of charge do not have the common features given in Eqs. (134) by accidental choice of examples. They are general in the sense that they hold for all electromagnetic plane waves. An electromagnetic plane wave consists of space- and time-dependent electric and magnetic fields $\underset{\sim}{E}(x,y,z,t)$ and $\underset{\sim}{B}(x,y,z,t)$ having the following properties:

1. There is a "propagation direction," which we take to be along $\hat{z}$. (the waves can be any combination of traveling or standing waves.)

2. None of the components of $\underset{\sim}{E}$ or $\underset{\sim}{B}$ depend on either of the transverse coordinates x and y . Thus we have

$$\underset{\sim}{E} = \hat{x}\, E_x(z,t) + \hat{y}\, E_y(z,t) + \hat{z}\, E_z(z,t) \qquad (159)$$

$$\underset{\sim}{B} = \hat{x}\, B_x(z,t) + \hat{y}\, B_y(z,t) + \hat{z}\, B_z(z,t) \quad . \qquad (160)$$

We are considering only linear, homogeneous, isotropic media. In fact we shall only consider the case of vacuum for simplicity. Everything we say holds throughout a given region of vacuum. Of course the fact that we have plane waves [waves of the form of Eqs. (159) and (160)] has something to do with where the waves came from, how they were produced, etc. We have considered in Sections 3.4 and 3.5 some examples in which we looked at the sources explicitly. Now we forget about the sources and just assume that the waves came from somewhere and have the form of Eqs. (159) and (160).

Now let us apply Maxwell's equations to Eqs. (159) and (160) First we use Gauss's law, that div $\underset{\sim}{E}$ is $4\pi\rho$. In vacuum ρ is zero. Also, since there are no x or y dependences of any of the components, the partial derivatives with respect to x and y give zero. Thus we have

$$\underset{\sim}{\nabla} \cdot \underset{\sim}{E} = \frac{\partial E_z(z,t)}{\partial z} = 0 \quad , \qquad (161)$$

which says that E_z is independent of z . That E_z is also independent of t can be seen by considering Maxwell's "displacement current" equation,

$$\frac{\partial \underset{\sim}{E}}{\partial t} = c\, \underset{\sim}{\nabla} \times \underset{\sim}{B} \quad . \qquad (162)$$

Take the z component of Eq. (162). The right hand side involves $\partial B_y/\partial x$ and $\partial B_x/\partial y$, both of which are zero. Thus $\partial E_z/\partial t$ is zero. We conclude that E_z is a constant. For simplicity we take the constant to be zero. In the problem of the radiating sheet we had a constant non-zero field $E_z = 2\pi\sigma$. We could have eliminated this constant field by adding a second

sheet of charge of density $-\sigma$ on top of the first, with the added sheet permanently at rest. Then the total field E_z would be zero, but the waves of transverse electric and magnetic fields would have been the same.

Similarly the fact that we have $\underset{\sim}{\nabla} \cdot \underset{\sim}{B} = 0$ tells us that $B_z(z,t)$ has no z dependence. That it has no time dependence either is seen by considering the z component of Faraday's law

$$\frac{\partial \underset{\sim}{B}}{\partial t} = - c \underset{\sim}{\nabla} \times \underset{\sim}{E} \ , \tag{163}$$

which gives $\partial B_z/\partial t$ to be zero. Thus, although there may be some steady magnetic fields due to big sheets of steady current somewhere, they have no space or time dependence and are not of present interest to us. We therefore take B_z to be not only constant but zero.

So far we have concluded that (aside from non-wave-like constant fields) electromagnetic plane waves are transverse waves. That is, the electric and magnetic fields are perpendicular to the direction of propagation $\hat{z}$.

We are left with E_x, E_y, B_x, and B_y, and the as-yet-unused x and y components of Eqs. (162) and (163):

$$\frac{1}{c} \frac{\partial E_x}{\partial t} = \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = - \frac{\partial B_y}{\partial z} \tag{164a}$$

$$\frac{1}{c} \frac{\partial B_y}{\partial t} = - \frac{\partial E_x}{\partial z} + \frac{\partial E_z}{\partial x} = - \frac{\partial E_x}{\partial z} \tag{164b}$$

$$\frac{1}{c} \frac{\partial E_y}{\partial t} = \frac{\partial B_x}{\partial z} - \frac{\partial B_z}{\partial x} = \frac{\partial B_x}{\partial z} \tag{165a}$$

$$\frac{1}{c} \frac{\partial B_x}{\partial t} = - \frac{\partial E_z}{\partial y} + \frac{\partial E_y}{\partial z} = \frac{\partial E_y}{\partial z} \tag{165b}$$

According to Eqs. (164), E_x and B_y are "coupled" by the two first order linear partial differential equations in which they occur. The time derivative of E_x is proportional to the (negative) space derivative of B_y, and the time derivative of B_y is proportional in the same way to the negative space derivative of E_x .

Equations (165) relate E_y and <u>minus</u> B_x in exactly the same way that Eqs. (164) relate E_x and B_y . Equations (164) and (165) are really the same equation, in the following sense: Suppose that only E_x and B_y are different from zero and that E_x and B_y are both positive. Now suppose that we leave the physical fields E_x and B_y alone but rotate the coordinate system about the z axis, so that $\underset{\sim}{E}$ now points along the (new) +y axis instead of along the (old) +x axis. Then we see that $\underset{\sim}{B}$, which used to be along the (old) +y axis is now along minus x . See Fig. 22.

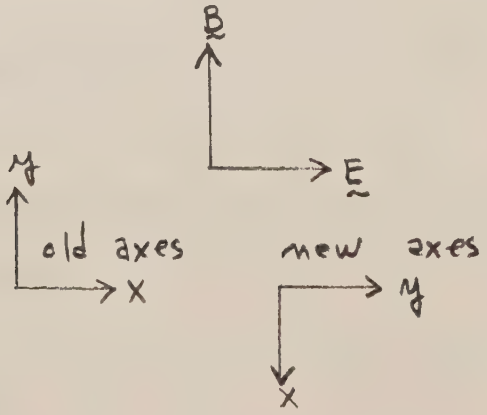


Fig. 3-22. A rotation of axes around z axis changes labels. E_x becomes E_y ; B_y becomes $-B_x$.

Thus we can confine ourselves to studying Eqs. (164). We know how to find field components that satisfy Eqs. (165) once we have found those that satisfy Eq. (164). That does not mean that we can always eliminate one set

of equations by a single rotation of the coordinate system. We might be able to do that at a certain time, but if the radiation is circularly polarized (see Chap. 4), the fields rotate in direction about the z axis and E_y is non-zero just as often as is E_x. What it does mean is that we can study a single linear "polarization state," i.e., a wave in which only E_x and B_y are non-zero, to find the relations between E_x and B_y for such a state. Then by a rotation we know the relations between E_y and $-B_x$ in the other possible linear polarization state with fields at 90 degrees to those of the first state. Finally, once we know the properties of both of these linear polarization states, we can superpose them with arbitrary amplitudes and phase constants to obtain the most general plane wave. Because we have <u>two</u> independent sets of coupled equations [Eqs. (164) and (165)],we have two <u>independent</u> states of polarization.

From now on we assume we have only the single linear polarization state corresponding to non-zero E_x and B_y, i.e., corresponding to Eqs. (164). It will be simplest if we consider first a pure harmonic traveling wave propagating in the +z direction. Then we will immediately see how to get the equivalent result for a pure harmonic wave propagating in the -z direction. A superposition of these with arbitrary amplitudes and phase constants is then the most general solution for a given frequency and includes pure standing waves as a special case. It is better for our understanding, however, if we consider a pure standing wave separately, as we shall.

<u>Traveling harmonic wave.</u> Assume that E_x is given by

$$E_x = A \cos(\omega t - kz) , \qquad (166)$$

so that

$$\frac{\partial E_x}{\partial t} = -\omega A \sin(\omega t - kz) \quad ,$$

then Eq. (164a) gives

$$\frac{\partial B_y}{\partial z} = -\frac{1}{c}\frac{\partial E_x}{\partial t} = \frac{\omega}{c} A \sin(\omega t - kz)$$

$$= k A \sin(\omega t - kz)$$

$$= \frac{\partial}{\partial z} A \cos(\omega t - kz)$$

$$= \frac{\partial E_x}{\partial z} \quad . \tag{167}$$

Similarly Eq. (164b) gives

$$\frac{\partial B_y}{\partial t} = -c\frac{\partial E_x}{\partial z} = -ck A \sin(\omega t - kz)$$

$$= -\omega A \sin(\omega t - kz)$$

$$= \frac{\partial}{\partial t} A \cos(\omega t - kz)$$

$$= \frac{\partial E_x}{\partial t} \quad . \tag{168}$$

Thus we see that B_y and E_x are equal, aside from uninteresting additive constants.

If we consider a harmonic traveling wave propagating in the $-z$ direction we find that B_y is the negative of E_x, as you can easily see by replacing k with $-k$ in the above equations. Both directions of propagation are included in the summarizing statements

$$\text{traveling wave:} \quad \begin{cases} |\underset{\sim}{E}(z,t)| = |\underset{\sim}{B}(z,t)| \\ \hat{\underset{\sim}{E}} \cdot \hat{\underset{\sim}{B}} = 0 \quad . \\ \hat{\underset{\sim}{E}} \times \hat{\underset{\sim}{B}} = \hat{\underset{\sim}{v}} \end{cases} \tag{169}$$

Standing harmonic wave. Assume that E_x is given by

$$E_x(z,t) = A \cos \omega t \cos kz \quad . \tag{170}$$

$$\frac{\partial E_x}{\partial t} = -\omega A \sin \omega t \cos kz$$

Then Eq. (164a) gives

$$\frac{\partial B_y}{\partial z} = -\frac{1}{c}\frac{\partial E_x}{\partial t} = \frac{\omega}{c} A \sin \omega t \cos kz = \frac{\partial}{\partial z}[A \sin \omega t \sin kz] \quad . \tag{171}$$

As far as the z dependence is concerned, this is satisfied by

$$B_y(z,t) = A \sin \omega t \sin kz \quad . \tag{172}$$

If you now check to see whether this B_y satisfies Eq. (164b), i.e., has the right time derivative, you will find that it does.

Thus we see from Eqs. (170) and (172) that in an electromagnetic standing wave, $\underset{\sim}{E}$ and $\underset{\sim}{B}$ are perpendicular to one another and to $\hat{z}$, have the same amplitude, and are 90 degrees out of phase in space and 90 degrees out of phase in time. This behavior is reminiscent of that of the pressure and velocity in a standing sound wave or of the transverse tension and velocity in a standing wave on a string.

General energy relations. Let us assume that the energy density is given by

$$\text{energy density} = \frac{1}{8\pi}\left[E_x^2 + B_y^2\right] \quad , \tag{173}$$

and see if we can find the appropriate expression for the energy flux vector.

(This expression agrees with Vol. 2, pp. 102 and 256, for the static cases and can be shown to hold generally.) The energy $W(z,t)$ in a volume element having area A perpendicular to the z axis and thickness Δz along the z axis is given by the energy density times $A\,\Delta\,z$:

$$W(z,t) = A\Delta z \frac{1}{8\pi}\left[E_x^2 + B_y^2\right] . \tag{174}$$

Let us find the time rate of change of this energy. Differentiating with respect to t we obtain

$$\frac{\partial W(z,t)}{\partial t} = \frac{A}{4\pi}\Delta z\left[E_x\frac{\partial E_x}{\partial t} + B_y\frac{\partial B_y}{\partial t}\right] .$$

This looks terrible. But now use Eqs. (164a) and (164b):

$$\frac{\partial W(z,t)}{\partial t} = -\frac{Ac}{4\pi}\Delta z\left[E_x\frac{\partial B_y}{\partial z} + B_y\frac{\partial E_x}{\partial z}\right]$$

$$= -\frac{Ac}{4\pi}\Delta z\frac{\partial}{\partial z}\left[E_xB_y\right] . \tag{175}$$

This looks better. We have used up all of our equations so we had better try to understand Eq. (175). Let us look at the Taylor's series for E_xB_y expanded at z,t, with t held constant:

$$E_x(z+\Delta z,t)B_y(z+\Delta z,t) = E_x(z,t)B_y(z,t) + \Delta z\frac{\partial(E_xB_y)}{\partial z} + \ldots$$

We see that the rate of change of energy in the volume $A\,\Delta\,z$ is given by

$$\frac{\partial W(z,t)}{\partial t} = Ac\frac{\left[(E_xB_y)_{z,t} - (E_xB_y)_{z+\Delta z,t}\right]}{4\pi} .$$

Thus the rate of change of energy in this volume element is the value of a certain quantity (the quantity being $cA\,E_xB_y/4\pi$) at the point z,t <u>minus</u> the value of the same quantity at $z+\Delta z,t$. It is clear that the quantity

must be the energy flow rate into Δz from the left at z and out of Δz to the right at $z + \Delta z$. Thus the flux vector $S_z(z,t)$, defined as the energy flow in the $+z$ direction, is given (in erg/cm^2 sec) by

$$S_z(z,t) = \frac{c}{4\pi} E_x(z,t)\, B_y(z,t) \quad . \tag{176}$$

Equation (176) holds in general for any combination of traveling or standing waves. We can put it in a coordinate-invariant form by writing

$$\underset{\sim}{S} = \frac{c}{4\pi} \underset{\sim}{E} \times \underset{\sim}{B} \quad .$$

There is one slight omission in our derivation. We called the expression of Eq. (173) by the name "energy density" and then proved that Eq. (176) gives the energy flux. But our derivation did not really show that the $1/8\pi$, for example, is necessary. In order to get all the proportionality constants right, one must not only find something that is conserved (i.e., flows without loss); one must also find how it is coupled to objects that do work. We have already done that in our examples that included the sources, so that we felt free to remember the correct proportionality constant in the energy density.

Characteristic impedance of vacuum. Let us find the characteristic impedance experienced by our plane wave traveling in vacuum in the $+z$ direction. Suppose $\underset{\sim}{E}$ is along x. The rate at which energy flows across an area A of edge lengths L_x and L_y is the product of the energy flux vector S_z given by Eq. (176) times the area $L_x L_y$:

$$P = S_z L_x L_y \tag{177}$$

This power P can be completely absorbed if the plane wave is terminated by a sheet consisting of a resistive impedance Z such that the potential

difference $V = E_x L_x$ produces a current $I = V/Z$ such that the rate at which the impedance Z absorbs energy, VI, is equal to P. Thus for perfect termination we need

$$VI = \frac{V^2}{Z} = \frac{E_x^2 L_x^2}{Z} = P = S_z L_x L_y \quad .$$

Thus the part of the terminating sheet that occupies the area $L_x L_y$ has a resistance Z for current flow in the x direction given by

$$Z = (E_x^2/S_z)\,(L_x/L_y) \quad . \tag{178}$$

Now we use Eq. (176) for S_z, and the fact that B_y equals E_x in a traveling plane wave, to obtain from Eq. (178)

$$Z = \frac{4\pi}{c}\,(L_x/L_y) \quad . \tag{179}$$

We can take the area $L_x L_y$ to be a square, with $L_x = L_y$. Then the "impedance per square" given by Eq. (179) is independent of the size of the square. We convert to MKS units by recalling that $1/c$ in statohms is 30 ohms. Thus the characteristic impedance Z experienced by plane electromagnetic waves in vacuum is given by

$$\begin{aligned} Z\ (\text{vacuum}) &= 120\ \pi \text{ ohms per square} \\ &= 377 \text{ ohms per square} \quad . \end{aligned} \tag{180}$$

One can make a sheet of material that will terminate microwaves without reflection by starting with cloth that has impedance greater than $377\ (L_x/L_y)$ ohms resistance for current flow in the x direction (for a potential applied uniformly across the $+x$ and $-x$ edges of a rectangular sheet of cloth) and then painting the cloth with "aquadag" (a suspension of carbon in water) until the resistance is decreased to $377\ (L_x/L_y)$. The resulting material is sometimes called "spacecloth."

*3.7 Reflection at a reactive load.

In Chapter 1 we found by studying a dispersion relation that radio waves are totally reflected back to earth from the ionosphere when their frequency is less than the cutoff frequency, which is that of the free oscillations of the electrons of the plasma. We also mentioned that one could use essentially the same dispersion relation to understand metallic reflection of visible light. In Chapter 2 we found that when a harmonic wave of given ω has a negative value of k^2, the solutions are given by cos ωt times $e^{-|k|z}$ or $e^{+|k|z}$. If the reactive region (region with negative k^2) extends to $+\infty$, only the solution with $e^{-|k|z}$ is present.

We wish to find what happens when a harmonic traveling electromagnetic wave propagating in the $+z$ direction is incident on a reactive medium extending from $z = 0$ to $+\infty$. The incident electric field in medium 1 (to the left of $z = 0$) is given by

$$E_{inc} = A \cos (k_1 z - \omega t) \quad . \tag{181}$$

We wish to find the form of the transmitted and reflected waves. The transmitted wave will have the form

$$E_{trans} = B \cos (\omega t + \phi)\, e^{-|k_2|z} \quad , \tag{182}$$

where $|k_2|$ is given by whatever dispersion relation holds, and B and ϕ are unknowns.

Since we expect phase shifts, i.e., ϕ in Eq. (182) and a similar one in the reflected wave, it will be easier if we use complex numbers. Then E_{inc} and E_{trans} are taken to be the real parts of the complex quantities

$$\psi_{inc}(z,t) = A e^{-i\omega t}\, e^{i k_1 z} \tag{183}$$

$$\psi_{trans}(z,t) = B\, e^{-i\phi}\, e^{-i\omega t}\, e^{i k_2 z} = B\, e^{-i\phi}\, e^{-i\omega t}\, e^{-|k_2|z} \tag{184}$$

According to Eq. (184), for k_2^2 negative k_2 is given by the positive square root of -1 times $|k_2|$:

$$k_2 = +i|k_2| \quad . \tag{185}$$

According to Eq. (140), the reflection coefficient for the electric field (if $\mu_1 = \mu_2 = 1$) is given by

$$R_E = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{Z_1^{-1} - Z_2^{-1}}{Z_1^{-1} + Z_2^{-1}} = \frac{k_1 - k_2}{k_1 + k_2} \quad . \tag{186}$$

This relation holds when medium 2 is a reactive medium as well as when it is a dispersive one. For the present case of a reactive medium, Eqs. (185) and (186) give

$$R_E = \frac{k_1 - i|k_2|}{k_1 + i|k_2|} \tag{187}$$

Notice that in the limit of $|k_2|$ huge compared with k_1, R_E is essentially -1 . If $|k_2|$ is huge compared with k_1, then the exponential decay distance $|k_2|^{-1}$ is small compared with the wavelength $2\pi k_1^{-1}$. (We note that according to Eq. (182) the transmitted wave "decays" by a factor $e = 2.718 \ldots$ for each distance $|k_2|^{-1}$ that the wave advances in medium 2.) That is commonly the situation encountered in metallic reflection, for example, reflection of visible light from a metal surface. In studying electrostatic fields one gets used to saying that the component of electric field tangential to the surface of a conductor is zero at the surface of the conductor. But that is only true if the electrons have plenty of time to move so as to cancel the applied field. More generally the tangential electric (and magnetic) fields fall off according to $e^{-|k_2|z}$ in the conductor.

In the above limiting case where R_E is -1 , the reflected wave has the same amplitude as the incident wave. Thus the reflected energy flux is the same as the incident flux. Does that condition hold only in that extreme

limit? No! It must be true for any value of $|k_2|/k_1$, since the transmitted wave eventually becomes as small as you wish, so that there can be no net transmitted flux at $z = +\infty$. Thus all of the incident flux must turn around and come back out after some time delay. That can be so only if the magnitude of R_E is unity. That this is true is easily seen:

$$|R_E|^2 = R_E^* R_E = \frac{k_1 + i|k_2|}{k_1 - i|k_2|} \cdot \frac{k_1 - i|k_2|}{k_1 + i|k_2|} = 1 \quad . \tag{188}$$

Since we are interested in the limiting case where R_E is approximately -1 , we shall make use of Eq. (188) to write R_E in the exact form

$$R_E = -e^{i\delta} = -\cos\delta - i \sin\delta \tag{189}$$

The quantities cos δ and sin δ are expressed in terms of k_1 and k_2 by multiplying numerator and denominator of Eq. (187) by $k_1 - i|k_2|$, collecting real and imaginary terms, and comparing with Eq. (194):

$$\cos\delta = \frac{|k_2|^2 - k_1^2}{|k_2|^2 + k_1^2} \quad , \quad \sin\delta = \frac{2|k_2|\ k_1}{|k_2|^2 + k_1^2} \quad . \tag{190}$$

The reflected wave has the complex wave function

$$\psi_{ref} = R_E A\, e^{-i\omega t}\, e^{-ik_1 z} = -A\, e^{-i(\omega t - \delta)}\, e^{-ik_1 z} \quad . \tag{191}$$

The reflected electric field is given by the real part of Eq. (191):

$$E_{ref} = -A \cos(k_1 z + \omega t - \delta) \quad . \tag{192}$$

Notice that positive δ implies a time delay for the reflected wave relative to the incident one. Thus we could write

$$\omega t - \delta = \omega(t - t_{delay}) \quad . \tag{193}$$

[We must be careful with such an interpretation because of the fact that with steady harmonic oscillations one can never tell which oscillation at one place is due to a certain one at another. The harmonic waves travel at the phase velocity. If you put a little "pip" on one wave crest at one place so that it can be "recognized" downstream, the pip will travel with the group velocity and will not stay on the wavelet that you pipped! Then you have the problem of trying to figure out whether it may have shifted by one or more wavelengths. However if the distance downstream is small enough, we can assume that that has not happened. Similarly in the present case, if the penetration distance is small compared with a wavelength, we shall interpret δ as due to a small delay of much less than one cycle of phase, rather than a large "advance" of almost an entire cycle of phase.] Let us see if we can guess t_{delay} from a crude intuitive picture. The mean penetration distance into the reactive medium is the exponential decay length $|k_2|^{-1}$. We might expect that when the incident traveling wave (traveling at phase velocity $v_1 = \omega/k_1$) suddenly finds itself in the reactive medium, it goes "down and back" a total distance $2\,|k_2|^{-1}$ at its old phase velocity. Thus we might guess

$$t_{delay} = \frac{2|k_2|^{-1}}{v_1} = \frac{2\,k_1}{\omega|k_2|} \tag{194}$$

For $|k_2| \gg k_1$, this guess is correct, as we see by comparing Eqs. (193) and (194) with the expression for $\sin\delta$ in Eq. (190) and recalling that for δ small we can write $\sin\delta \approx \delta$.

The total wave function $E_1(z,t)$ in the dispersive medium (medium 1) is the superposition of the incident wave and the reflected wave. According to Eqs. (181) and (192) we have

$$\begin{aligned} E_1(z,t) &= A\cos(k_1 z - \omega t) - A\cos(k_1 z + \omega t - \delta) \\ &= 2A\cos(\omega t - \tfrac{1}{2}\delta)\sin(k_1 z - \tfrac{1}{2}\delta)\,. \end{aligned} \tag{195}$$

The second equality in Eq. (195) is most easily derived using the complex functions.

At the boundary at $z = 0$, the transmitted electric field in medium 2 is equal to the total (incident plus reflected) field in medium 1 if we assume that there is no layer of surface charge there to give a discontinuity in the electric field. Thus we can write down the transmitted field by inspection of Eq.(195) [to get the value at $z = 0$], using the form given by Eq.(182. We have

$$E_{trans} = -2A \sin \tfrac{1}{2}\delta \cos(\omega t - \tfrac{1}{2}\delta)\, e^{-|k_2|z}, \qquad (196)$$

which has the form of Eq. (182) and reduces to Eq. (195) at $z = 0$.

In Fig. 23 we plot Eqs. (195) and (196) at an instant when $\cos(\omega t - \frac{1}{2}\delta)$ is -1.

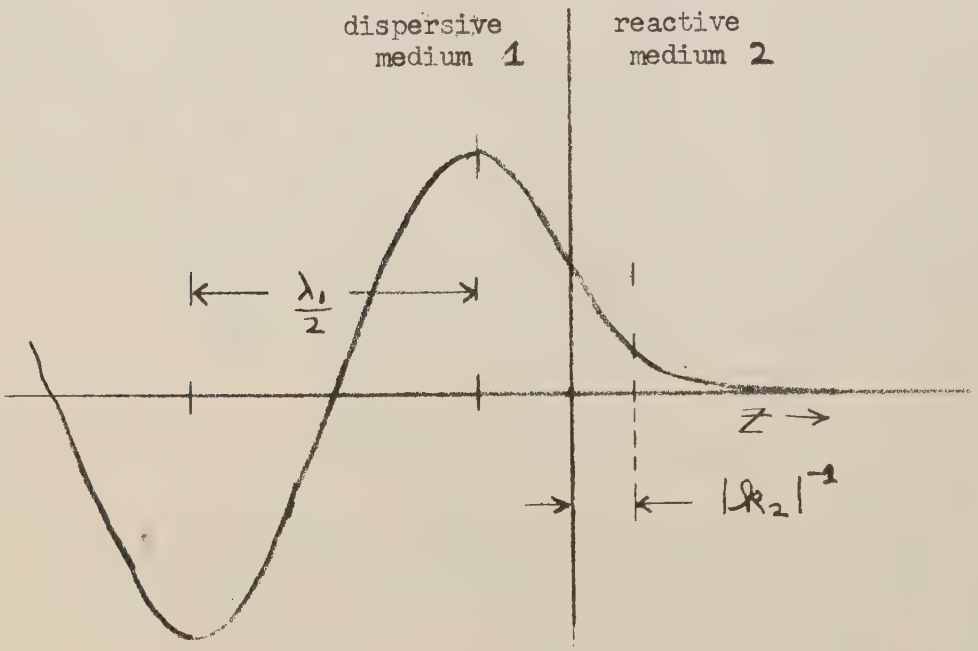


Fig. 3-23. Reflection at boundary between dispersive and reactive media. For this example $\frac{1}{2}\delta$ is $\pi/4$.

Now we shall give some examples:

Radio waves in the earth's ionosphere. We discussed this in Chapter 1, starting on page 61.

Metallic reflection of visible light from silver. As an approximation we assume that, since the valence of silver is one, each silver atom in a piece of silver contributes one "free" electron. We shall assume that the free electrons behave just as in the ionosphere. Thus we have [Eq. (110), Chap. 1, p. 62]

$$k_2^2 = \frac{\omega^2}{c^2} - \frac{\omega_p^2}{c^2} = k_1^2 - \frac{\omega_p^2}{c^2} ; \quad \omega_p^2 = \frac{4\pi N e^2}{m} . \quad (197)$$

Using the density (10.5 gm/cm^3), atomic weight (107.9 gm), and valence (1) of silver, and Avogadro's number (6.0×10^{23}), we find the number of "free" electrons per unit volume in solid silver to be $N \approx 6.0 \times 10^{22}$ cm^{-3} . The plasma oscillation frequency is then 2.2×10^{15} cps. For visible light of wavelength 5500 A^o(green) the frequency is 5.4×10^{14} cps. This is slower than the plasma oscillation frequency by a factor $\approx \frac{1}{4}$. Thus we have $R_E \approx -1$. The exponential attenuation distance $|k_2|^{-1}$ is essentially c/ω_p , which is about 1.4×10^{-5} cm.

From what we have said, you might expect that for frequencies lower than those of visible light, for example for microwave frequencies, silver (or copper, or any other conductor) would continue to act like a purely reactive medium. But that is not so. In our derivation of Eq. (197) we neglected energy transfers from the electrons due to collisions. When the frequency becomes low enough, the electrons in a solid conductor lose energy by collisions with lattice ions, etc. In fact they reach a constant velocity, the "terminal velocity," which means the conductor acts like a purely resistive medium. That is the way metals behave with constant voltages applied to them, and that is also the way they behave when microwaves are incident on them. The frequency of transition between resistive and reactive behavior for silver turns out to be about 2.5×10^{14} cps. Slower than that, silver is mostly

resistive. Faster than that, silver is mostly reactive. For still higher frequencies, above the cutoff frequency, it becomes transparent. See problem 23 for a detailed discussion.

Rectangular wave guide. Consider the parallel-plate transmission line filled with vacuum discussed in Sec. 3.4. We show in Fig. 24 a modified version of Fig. 14 from that section.

Fig. 3-24. Parallel-plate transmission line. Traveling wave propagates in z direction. Arrows denote electric field at instant shown. Plus and minus signs denote surface charge. The plates extend infinitely far in the lateral (plus or minus y) directions, as denoted by "etc."

The dispersion relation for this system is just that for electromagnetic waves in vacuum:

$$k^2 = \frac{\omega^2}{c^2} \quad . \tag{198}$$

The waves are traveling waves of the form

$$E_x(z,t) = A \cos (\omega t - kz) \quad , \tag{199}$$

satisfying the one-dimensional wave equation

$$\frac{\partial^2 E_x(z,t)}{\partial t^2} = c^2 \frac{\partial^2 E_x(z,t)}{\partial z^2} \quad , \tag{200}$$

with the magnetic field $B_y(z,t)$ determined by the known relations between electric and magnetic fields in a traveling wave.

Suppose that we insert conducting "sideplates" at two values of y along the transmission line. The tangential electric field must go to zero at these conductors (which have no surface charge). The electric field can continue to be along plus or minus x, as it was for the transmission line, but now we have eliminated the possibility of having no y dependence for E_x. Thus we now have $E_x(y,z,t)$. The wave equation is now the two-dimensional classical wave equation,

$$\frac{\partial^2 E_x(y,z,t)}{\partial t^2} = v^2 \frac{\partial^2 E_x(y,z,t)}{\partial z^2} + v^2 \frac{\partial^2 E_x(y,z,t)}{\partial y^2} \quad , \tag{201}$$

with $v^2 = c^2$. Since E_x must remain zero at the side plates, E_x must have a standing-wave character with respect to the y direction. We know that as far as the wave equation is concerned, it doesn't care whether its solutions are traveling or standing waves. In the present case the solution that fits the boundary conditions is traveling in the z direction, but is a standing wave with respect to the y direction. Thus Eq. (201) has solutions

$$E_x(y,z,t) = A \sin(k_y y) \cos(\omega t - k_z z) \quad , \tag{202}$$

which (for fixed z) is a standing wave in the y direction and (for fixed y) is a traveling wave traveling in the $+z$ direction.

The parallel-plate transmission line with added conducting side plates is called a rectangular wave guide. The modes we have described are called TE (transverse electric) modes. The lowest TE mode is one where there is just one half cycle of $\sin k_y y$ between the side plates. This is shown in Fig. 25.

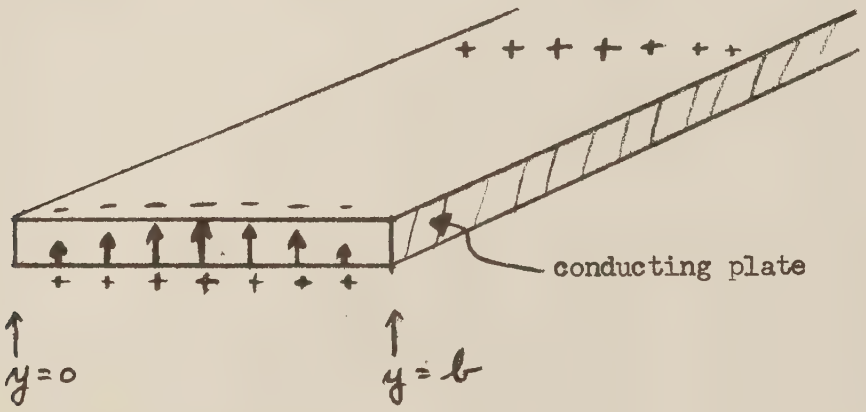


Fig. 3-25. Rectangular wave guide. Conducting side plates have been added to the transmission line of Fig. 24. The arrows give the electric field. The electric field is always zero at the side plates.

We have already satisfied the boundary condition at $y = 0$ by writing $\sin k_y y$ instead of $\cos k_y y$. To satisfy the boundary condition at $y = b$ in the lowest mode, we need $k_y b = \pi$, corresponding to one half cycle between $y = 0$ and b.

The dispersion relation for this (TE) mode is obtained by inserting Eq. (202) into the wave equation, Eq. (201), to yield

$$\frac{\omega^2}{c^2} = k_z^2 + k_y^2 = k_z^2 + (\pi/b)^2 \quad , \tag{203}$$

i.e.,

$$k_z^2 = \frac{\omega^2}{c^2} - (\pi/b)^2 \quad . \tag{204}$$

Now let us look at Eq. (204) as a function of frequency. For very high frequency, k_z is essentially ω/c. Then the phase velocity is c, and the wavelength along the z direction is the wavelength for free plane waves in vacuum. (The waves of the form of Eq. (202) are not plane waves. Why not?) As ω decreases, k_z decreases even more (proportionally), so that the phase velocity increases (it starts at c for infinite ω). For ω less than the cutoff frequency, $\omega_{c.o.}$, the wave guide becomes purely reactive, and the waves are exponentially attenuated with distance. The cutoff frequency is given by

$$\omega_{c.o.} = c\,\pi/b \quad . \tag{205}$$

Another way to think of the transition from dispersive to reactive behavior is to think of ω as held fixed while b is varied. When b is very large, the wave propagates as a plane wave. When b gets so small that it is one half wavelength of the plane wave having frequency ω in vacuum, we are at cutoff. For b shorter than $b_{c.o.}$ the medium is reactive. Thus if a wave guide with large value of b (greater than $b_{c.o.}$) is joined to one with a small value of b (less than $b_{c.o.}$), we get total reflection of the wave and exponential attenuation in the second guide, with the field in the second guide given by

$$E_x(y,z,t) = A \sin(k_y y) \cos \omega t \; e^{-|k_z| z} \quad . \tag{206}$$

Barrier penetration. If the reactive guide (denoted by medium 2) extends from $z = 0$ to $z = L$ and is joined at $z = L$ by another dispersive medium (denoted by medium 3), the transmitted wave incident on the third region is given by Eq. (206) with $z = L$. The amplitude of the

wave transmitted into the new dispersive region will be proportional to Eq. (206) evaluated at $z = L$. Thus we will have

$$E_x(\text{third region}) = B \sin(k_y y)\, e^{-|k_{2z}|L} \cos(\phi + \omega t - k_{3z}) \,. \quad (207)$$

The important factor is the exponential factor. If L is not too huge we get a non-negligible penetration of the wave through the barrier and into the dispersive region 3.

*Total internal reflection of light in glass. Suppose we have a lot of glass from $z = -\infty$ to $z = 0$. At the plane $z = 0$ the glass ends, and vacuum begins and extends to $z = +\infty$. You might think that the vacuum would always act as a dispersive medium, as it does for plane waves. But you have already seen in the example involving the rectangular wave guide that when we do not have plane waves, i.e., when E_x varies along the y axis as well as along z, the axis of propagation, the wave guide becomes reactive under certain conditions (width too narrow, or what amounts to the same thing, frequency too low), even though it still has nothing in it but vacuum. A similar thing happens for light incident from glass to air if the angle of incidence becomes too large -- i.e., as the light gets too near to grazing incidence. This has a great practical importance. Many optical instruments make use of total internal reflection in order to get 100% reflection. An example is shown in Fig. 26. You might ask, why not just use a silvered piece of glass. The answer is that for visible frequencies silver is not completely reactive. It has some resistive behavior (see problem 23). Also the

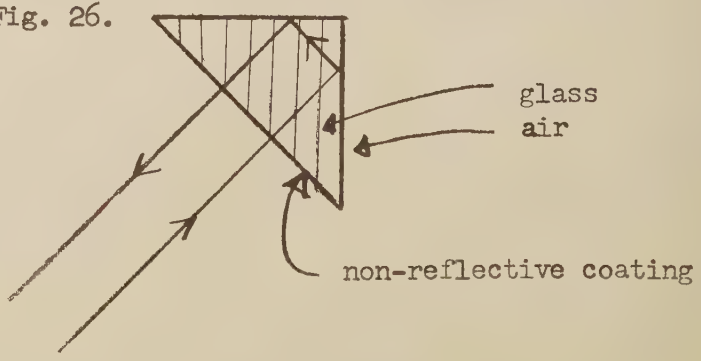


Fig. 3-26. Glass prism used to reflect light without loss.

silver tarnishes and gets even more resistive. (The fact that tarnished silver doesn't look so silvery any more proves that some visible light is being absorbed, i.e., that the medium has resistive properties. A surface that is completely black is purely resistive, since it must match impedances with the incident waves -- otherwise there would be reflections.)

Now let us see how it works. The light waves satisfy the wave equation in each medium, glass and air. (We are considering a single frequency ω.) The boundary between the glass and air is at $z = 0$. The propagation vector $\underset{\sim}{k}_1$ of the incident wave has a component k_z along $\hat{z}$ and a component k_y along $\hat{y}$. Thus we have a two dimensional problem (somewhat as we did for the wave guide in the TE mode). The geometry is shown in Fig. 27.

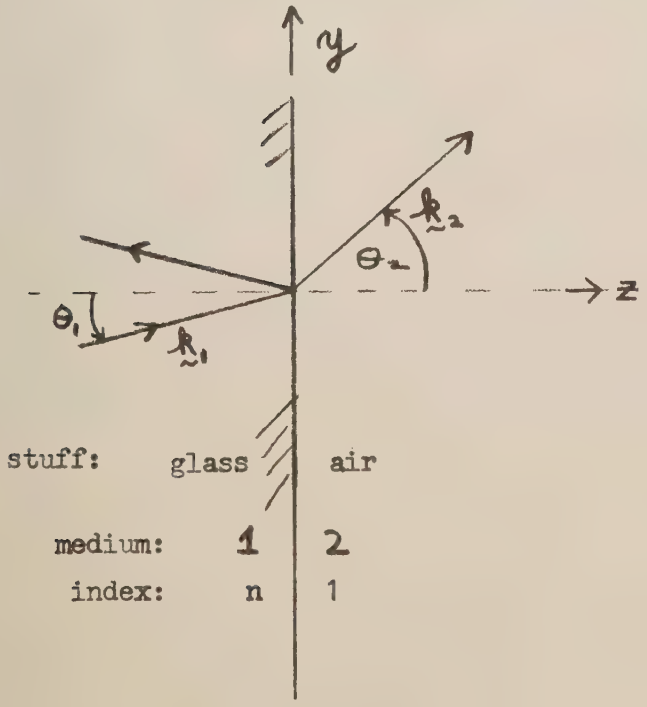


Fig. 3-27. Relations between propagation vectors in glass and air.

In either the glass or the air we have waves that satisfy the two dimensional wave equation, Eq. (201), with $v = c$ in air and $v = c/n$ in glass. (We consider air to be indistinguishable from vacuum.) The dispersion relation is given by

$$\frac{\omega^2}{c^2} = k_y^2 + k_z^2 \quad , \qquad (208)$$

where we must now find k_y and k_z in each medium. First, from Fig. 27 we see that the y component of $\underset{\sim}{k}_1$ is just the magnitude of $\underset{\sim}{k}_1$ times the sine of the angle of incidence θ_1 :

$$k_{1y} = k_1 \sin\theta_1 = n\frac{\omega}{c}\sin\theta_1 . \qquad (209)$$

Next, we claim that k_{2y} must equal k_{1y} . This is because the meaning of k_{1y} is 2π times the number of wave crests in medium 1 per unit length along $\hat{y}$. Similarly k_{2y} is 2π times the number of wave crests per unit length along $\hat{y}$ in medium 2. But as you travel along the y axis at $z = 0$, the number of crests just inside the glass that you pass must be the same as the number just outside the glass in the air. You cannot "lose" any crests per unit length along $\hat{y}$ in going from the glass to the air. Thus

$$k_{2y} = k_{1y} \qquad (210)$$

This relation can be written (by inspection of Fig. 27) as

$$k_1 \sin\theta_1 = k_2 \sin\theta_2 . \qquad (211)$$

Since k is proportional to the index of refraction, this becomes

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 , \qquad (212)$$

which is just $\sin\theta_2$ in our example where we have $n_2 = 1$ (to a sufficiently good approximation). Equation (212) is called Snell's law. It tells you how to find θ_2 if you know θ_1, or vice versa (assuming you know n_1 and n_2). Snell's law is the fundamental relationship for what is called geometrical optics. Using Snell's law you can find the path of any "ray of light" as it traces its way through an optical system having successive glass and air

interfaces. If you are designing an optical instrument, you can find where a ray would go (or a lot of rays having various directions and starting points). Then if you don't like where the rays go, you change something -- the kind of glass used in some places, the angles of the surfaces in others, etc. -- then put the new geometry and Snell's law into your IBM ray-tracing program and try again. A microsecond later, if you don't like the new results you try again, etc., etc. That is the way compound lenses are designed today -- by brute calculating force and Snell's law. (They used to be designed by people who understood how to clean up one term after another in certain Taylor's series expansions.)

The relation $k_{2y} = k_{1y}$, Eq. (210), is more fundamental than Snell's law. It <u>always</u> holds, whereas you may note that for our case, where Eq. (212) is given by

$$n \sin \theta_1 = \sin \theta_2 \ , \tag{213}$$

there is no solution for θ_2 if $\sin \theta_1$ is larger than n^{-1} (n is about 1.5 for glass). When $\sin \theta_1$ is n^{-1}, θ_1 is said to be at the <u>critical angle</u> θ_{crit}:

$$n \sin \theta_{crit} = 1 \ . \tag{214}$$

For $\theta_1 > \theta_{crit}$, Snell's law says that $\sin \theta_2$ is greater than 1. That has no meaning. According to the approximation called geometrical optics, one just says that there is no transmitted ray for $\theta_1 > \theta_{crit}$, and the incident ray is totally internally reflected.

But we know that we are dealing with waves, not "rays" (whatever "rays" are). Therefore when we encounter total reflection, we suspect that a reactive medium is involved. And so it is. Let us find k_{2z}^2. According to Eqs. (208), (210), and (209) we can write

$$k_{2z}^2 = \frac{\omega^2}{c^2} - k_{2y}^2 = \frac{\omega^2}{c^2} - k_{1y}^2$$

$$= \frac{\omega^2}{c^2} - n^2 \frac{\omega^2}{c^2} \sin \theta_1$$

$$= \frac{\omega^2}{c^2} [1 - n^2 \sin^2 \theta_1] \quad . \qquad (215)$$

Equation (215) holds for all incident angles θ_1. For θ_1 less than the critical angle, k_{2z}^2 is positive, and there is a transmitted traveling wave with k_{2y} given by $k_1 \sin \theta_1$ and k_{2z} given by Eq. (215). For θ_1 greater than the critical angle, the vacuum region is purely reactive, k_{2z}^2 is negative, and the transmitted electric field is given by

$$E_2 = B \cos (k_{2y} y - \omega t) \, e^{-|k_{2z}| z} \quad . \qquad (216)$$

In medium 2 the energy density falls off in the z direction like the square of the fields, i.e., like

$$e^{-2|k_{2z}| z} \quad .$$

A beautiful and ingenious experiment which verifies this exponential attenuation of the energy density has been performed by D. D. Coon, a graduate student (as of spring 1965) at Princeton.* Although his experiment is inherently quantum mechanical, it verifies this result of classical optics. This is but one of many results of classical optics which are retained in quantum mechanics. (Classical optics is what we have been studying whenever we dealt with electromagnetic waves of those wavelengths which we call "light," instead of, say, "microwaves"). Coon set two prisms with a variable air gap so that light (from the green line of mercury) was incident through one prism on the

* D. D. Coon, American Journal of Physics, 34, 240 (1966).

gap at an angle beyond the critical angle. The light energy transmitted through the air gap into the second prism is proportional to the energy density at the surface of the second prism. Now, quantum mechanics tells us that light of frequency ω comes in indivisible units, called photons, each with energy exactly $\hbar\omega$. Thus for a given ω the energy density is proportional to the number of photons. Coon measured the energy density by counting transmitted photons as a function of the air gap and verified the predicted exponential dependence. His experiment was the first to verify it at wavelengths of less than 1 cm and the first to verify it at any wavelength by detecting individual photons.

Problems and Home Experiments, Chapter 3

1. Show that the potential energy density (P.E. density) for a wave on a stretched string is given by

$$\text{P.E. density} = \tfrac{1}{2} T \left[\frac{\partial \psi(z,t)}{\partial z}\right]^2 \qquad \text{(P1)}$$

independent of whether the waves are traveling or standing waves. Hint: Consider the work that it takes to stretch a string segment so that it covers the same amount of distance along z when its slope is no longer zero as it did at equilibrium.

2. Show that for a sound wave the general expression for the energy flux $S_z(z,t)$ given by Eq.(93) (in units of ergs/cm^2 sec) is consistent with the result for emission of a pure traveling wave by a transmitter, Eq. (78). See the discussion following Eq. (93).

3. Show that in a traveling sound wave the energy density is twice the kinetic energy density and hence that the potential and kinetic energy densities are equal.

4. This problem is related to the discussion of barely audible sounds, near Eq. (84).

(a) Calculate the maximum gauge pressure g_o (where $p - p_o = g_o \cos \omega t$) for the case of barely audible sound having $\langle I \rangle = 10^{-10} I_o$ [with I_o given by Eq. (81)].

(b) Suppose that a flea or microbe or something is standing on your eardrum. (Your head is on its side so the bug doesn't fall off.)

Suppose the density of the bug is that of water. Pick a number for the area of the eardrum (to an order of magnitude). Now comes the question: If the extension of the eardrum caused by the bug is the same as the amplitude of vibration of a barely audible sound, what are the approximate linear dimensions of the bug? (Take a cubical bug.)

5. Suppose you have a loudspeaker with 10 watts audio output. The vibrating part of the speaker is a piston of 3 in. diameter pushing on the air in a 3 in. diameter tube. The tube is gradually flared out in a "horn" and eventually matches the walls of a room with 10 ft. by 10 ft. cross section and with a rear wall that is perfectly terminated by various kinds of felt or "mazes." The geometry is shown in Fig. P1.

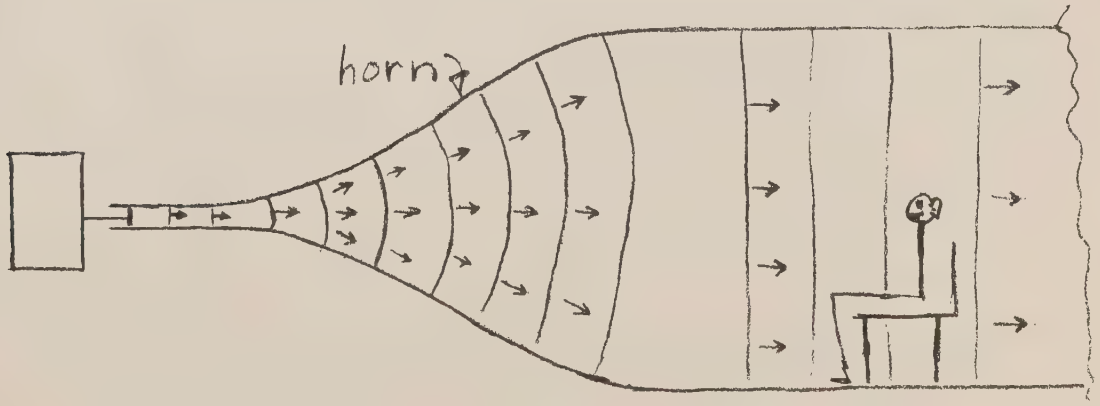


Fig. P1. Loudspeaker and impedance-matching horn.

The impedance per unit area at the piston is 42.8 gm/cm^2 sec. The impedance is this number multiplied by the area of the tube. The impedance of the room is larger than that of the tube by the ratio of the areas. If the tube were joined "suddenly" to the room, the mismatch (discontinuity) in impedances

would cause a reflection. By making a smooth transition over a distance of several wavelengths, i.e., by using a horn, one can "match impedances" between the speaker and the room and thus eliminate the reflections.

(a) For the numbers given above, what is the amplitude D_o of the speaker's piston vibrations, $D(t) = D_o \cos \omega t$, for $\nu = 50$ cps and for $\nu = 15{,}000$ cps.

(b) How loud is it at the listener's ear (in units of I_o)?

*6. Impedance matching for sound waves. Suppose that you want to match impedances between a tube of area A_1 and one of area A_2 by inserting an intermediate tube of area A_m and of length L_m between tubes 1 and 2, as shown in Fig. P2.

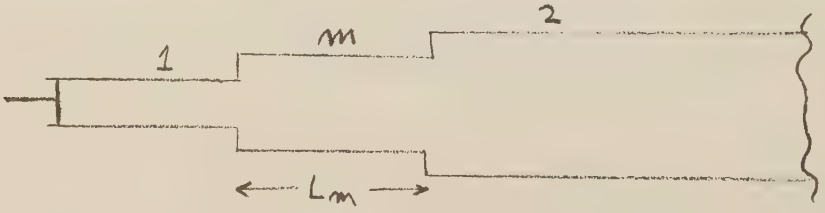


Fig. P2.

(a) We shall assume that you only want to eliminate reflections for a single frequency ν_o with wavelength λ_o. What should you choose for L_m and for the proportionality $Z_1 : Z_m : Z_2$? Hints: For simplicity consider only the limit where tube 2 is only slightly greater than tube 1. Then you can assume that the amplitude of the transmitted traveling wave incident on the second discontinuity is essentially that of the traveling wave incident on the first discontinuity, and you need not worry about "multiple reflections." Actually the answer you will then get is the right answer even in the limit of large changes of impedance. Answers:

$L_m = \frac{\lambda_o}{4}$; $z_m = \sqrt{Z_1 Z_2}$.

(b) Notice that at a place where the radius suddenly increases the air must flow around the corner. The conditions are thus not exactly those which we have studied, at least not until the air gets well into the tube. That means that the impedance is not going to be exactly proportional to the area of the tube, as it is when we have a flat piston pushing on a tube. Can you design a mechanical coupling system to go from one tube to the next such that the gas never flows radially, and thus such that our formulas for impedance hold exactly? Do not worry about practical considerations; we want an ideal solution.

(c) Using the "small-reflections" approximation of part (a) (so as not to worry about multiple reflections], you have found how to choose L_m and A_m so as to eliminate reflections for traveling waves of frequency ν_o incident from the left. Now calculate what happens at other frequencies. Derive a formula for the fractional reflected intensity as a function of frequency. Use the small reflections approximation. Naturally your answer should give zero reflected intensity at $\nu = \nu_o$. At what other frequencies will you get zero? What happens for frequencies in between?

7. Prove the statement made in the sentence that includes Eq. (94).

8. Given $\partial\psi(z,t)/\partial t$ for all z and t and $\partial\psi(z,t)/\partial z$ for all z and t, and given $\psi(0,t)$ for all t, show how one can find $\psi(z,t)$ for all z and t. [See the discussion following Eq. (94).] Hints: Start with the Taylor's series expansion

$$\psi(z,t) = \psi(0,t) + z\left[\frac{\partial\psi(z,t)}{\partial z}\right]_{z=0} + \frac{1}{2}z^2\left[\frac{\partial^2\psi(z,t)}{\partial z^2}\right]_{z=0}$$

$$+ \frac{z^3}{3!}\left[\frac{\partial^3\psi(z,t)}{\partial z^3}\right]_{z=0} + \cdots \qquad \text{(P2)}$$

and use a suitable generalization of the result of problem 7 (i.e., use the wave equation). To prove that you understand what to do, give the expressions needed in the Taylor's series up through the term in z^5. The word "etc." completes the proof.

9. (a) Explain why the kinetic energy density (K.E.D.) in a general sound wave is given by

$$\text{K.E.D.} = \frac{1}{2}\rho_o v^2(z,t) = \frac{1}{2}\rho_o\left[\frac{\partial\psi(z,t)}{\partial t}\right]^2 , \qquad \text{(P3)}$$

and <u>not</u> the expression you would get by replacing ρ_o by $\rho(z,t)$ in Eq. (P3).

(b) Use energy conservation, the fact that in a standing wave the energy goes back and forth from 100% kinetic to 100% potential, and Eq. (P3) for the kinetic energy density to show that for a general sound wave the potential energy density (P.E.D.) is given by

$$\text{P.E.D.} = \frac{1}{2}(\gamma p_o)^{-1}[p(z,t)-p_o]^2 = \frac{1}{2}\gamma p_o\left[\frac{\partial\psi(z,t)}{\partial z}\right]^2 . \qquad \text{(P4)}$$

You may notice that the energy densities for transverse waves on a string can be expressed in a manner analogous to that of Eqs. (P3) and (P4) if we write

$$\text{string velocity (transverse)} = v(z,t) = \frac{\partial\psi(z,t)}{\partial t} \qquad \text{(P5)}$$

$$\text{string tension}\left(\begin{matrix}\text{acting on string}\\ \text{to right of z}\end{matrix}\right) = T_x(z,t) = -T_o \frac{\partial \psi(z,t)}{\partial z} \quad . \qquad (P6)$$

Then we have

$$\text{K.E.D. (string)} = \tfrac{1}{2}\rho_o v^2(z,t) = \tfrac{1}{2}\rho_o \left[\frac{\partial \psi(z,t)}{\partial t}\right]^2 \qquad (P7)$$

$$\text{P.E.D. (string)} = \tfrac{1}{2}(T_o^{-1})\, T_x^{\,2}(z,t) = \tfrac{1}{2} T_o \left[\frac{\partial \psi(z,t)}{\partial z}\right]^2 \qquad (P8)$$

10. Using $v(z,t)$ and $T_x(z,t)$ for the transverse waves on a string as defined by Eqs. (P5) and (P6), show that the two coupled first order differential wave equations are given by

$$\rho_o \frac{\partial v(z,t)}{\partial t} = -\frac{\partial T_x(z,t)}{\partial z} \qquad (P9)$$

$$\frac{\partial T_x(z,t)}{\partial t} = -T_o \frac{\partial v(z,t)}{\partial z} \quad . \qquad (P10)$$

These two equations show how the transverse velocity changes are induced by gradients in transverse force and how changes in transverse force are induced by gradients in transverse velocity. Equations (P9) and P10) play the same role for the waves on a string as Eqs. (95) and 96) play for sound waves.

11. Whiskey-bottle resonator (Helmholtz resonator).

If you blow across the mouth of a jug or bottle you get a tone because you have excited the lowest mode. If you try to estimate what frequency to expect, assuming that the bottle acts like a uniform tube closed at one end so that the length from bottom to top is $\frac{1}{4}\lambda$, you will get a surprise. The tone is much lower than you guessed. Here is Helmholtz's approximation. It gives fairly good results. (Using an empty Jim Beam bottle, I predicted 110 cps and found 130 cps, according to my piano.) Assume that the air in the large volume V_o

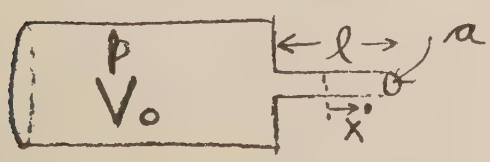


Fig. P3

acts like a spring. It is attached to a mass which is the air in the neck of the bottle. This mass is $\rho_o a\ell$, where ℓ is the length of the neck and a is its area. The approximation is to assume that all the motion is in the neck and that all the return force comes from pressure changes in V_o.

(a) Show that if x is the outwards displacement of fluid along the neck, and if all the return force F_x is given by $p - p_o$, where p is the pressure in V_o, then

$$F_x = -\frac{\gamma p_o a^2 x}{V_o} \qquad \text{(P11)}$$

(b) Show that this gives the single mode an angular frequency ω given by

$$\omega^2 = v^2(a/V_o\ell) , \qquad \text{(P12)}$$

where v is the velocity of sound. In using Eq. (P12) one must replace ℓ by the "effective" length of the neck, which is the actual length plus about 0.6 of a neck radius at each end. If the actual neck length is zero, the formula still works fairly well (the length ℓ then being completely due to "end corrections"). In that case one can use something like one of the rectangular parallelepiped cans that paint thinner comes in.

12. Use the same approximate model as was used in problem 11 to obtain a formula for the lowest resonant frequency of a uniform tube closed at one end. Assume that all the return force is at the open end and is given by $(p - p_o)\,A$, where A is the area of the tube. Let x be the motion of fluid at end of tube. Then the center of mass of the air moves a distance $\frac{1}{2}x$. The value of x gives the change in volume. Show that this approximation gives

$$\omega^2 = v^2\,(2A/V_oL_o) = v^2(2/L_o^2) , \qquad \text{(P13)}$$

where V_o is the volume and L_o is the length. (V_o is AL_o) This formula looks like Eq. (P12), because the derivation is similar. Of course the exact answer is given by taking L_o to be $\frac{1}{4}\lambda$. Thus the exact answer (except for an end correction on L_o) is

$$\omega^2 = v^2k^2 = v^2(2\pi/\lambda)^2 = v^2(\pi/2L_o)^2 . \qquad (P14)$$

Thus Eq. (P13) underestimates the frequency by a factor $2u2/\pi = 0.90$. The trouble was not with the physics (Newton's laws) but with our estimate of the average motion of the center of mass.

13. Show that the sound impedance at (or slightly outside of) the open end of a tube is zero and that at a closed end the impedance is infinite. Show that if R_{12} means the reflection coefficient for waves incident "from 1 to 2," and if R_p means the reflection coefficient for pressure and R_v that for velocity, then we have

$$R_{v,12} = -1 \text{ at a closed end, and } +1 \text{ at an open end}$$

$$R_{p,12} = -1 \text{ at an open end, and } +1 \text{ at a closed end,}$$

where in every case "from 1 to 2" means the incident wave is inside the tube and medium 2 is outside.

14. Suppose you take a deep breath of air and sing the note C 523. Now take a deep breath of helium and sing, keeping your laryngeal geometry the same as it was. What note will come out? Assume that the change of impedance will not disturb you, or if you wish, blow across a pop bottle that has been tuned to C 523 by partially filling it with water. You will find a table of frequencies in the handbook of chemistry and physics. Use the equal-temperament scale (one with equal ratios between successive chromatic notes -- the ratio is $2^{1/12}$) in which $A_4 = 440$ cps (the first A above middle C).

Answer: C 523.3 (air at STP) $\rightarrow$ 1536 cps $\approx G_6$ 1568 (helium at STP). This is higher by a factor of not quite 3, i.e., by not quite an octave plus a fifth.

*15. Suppose that a long tube filled with air at STP is separated by a membrane from a continuation of the tube filled with helium at STP. The membrane keeps the gases separated but moves freely. A traveling <u>compressional</u> pulse (region of excess pressure) travels down the air tube and is incident on the boundary (the membrane). Describe the reflected pressure pulse; i.e., is it a compression or a rarefaction? How does its maximum "height" compare with that of the incident pulse? Design an impedance matching section of tube of the same diameter such that for C 523 Hz (one Hz = 1 Hertz = 1 cps) the reflected wave will be zero. Use the air and helium that are presumably at your disposal, and choose the length wisely. Assume that γ_m and ρ_m for the mixture of air and helium that you use are given by linearly-weighted averages of the proportions of air and helium. The pressure is 1 atmosphere in all three sections. Hints: Don't look for a fine but messy formula; put in numbers early, not late in the

derivation.

Answer: You want $Z_m^2 = Z_1 Z_2$ and $L_m = (1/4)\lambda_m$. (Show these first.) Let f be the fraction of air (by volume) and 1-f the fraction of helium. The molecular weight of helium gas (which is monatomic) is 4 gm/mole; that of air is 29 gm/mole (a mixture of diatomic nitrogen having 28 and oxygen having 32 gm/mole); γ for air is 1.4; γ for He is 1.67. Show that to satisfy the requirement on Z_m you need f = 0.24. This gives $\gamma_m = 1.14\gamma$ (air) and $\rho_m = 0.35\ \rho$(air), and hence $v_m = 1.83$ v(air) (show this). For air at STP, v(air) is 3.32×10^4 cm/sec. Then for 523 Hz, λ(air) is 63.5 cm (show this); 1/4 λ is 15.9 cm. For $v_m = 1.83$ v, L_m must be 1.83(15.9) = 29.1 cm. One reason why it is bad practice to try to solve a problem like this with one grand and terrible formula is that you don't learn all the interesting values of the intermediate variables which you so grandly eliminate in order to answer what you thought was the question.

16. Derive Eq. (127) for B_ϕ for a coaxial cable. Then derive Eq. (128). Similarly derive Eqs. (129) and (130) for E_r and C_z.

*17. Impedance matching for light waves. (This problem is exactly analogous to problem 6.) Suppose you have light incident from medium 1 to medium 2. The reflection coefficient for the electric field in the light wave is given by Eq. (140) with $\mu_1 = \mu_2 = 1$, $R_E = (n_1 - n_2)/(n_1 + n_2)$. Suppose that for light at a given frequency ν_o you wish to completely eliminate the reflection. How can you do it? See problem 6! Answer: put a layer (called layer number 3) between medium 1 and medium 2 such that the index of refraction n_3 is the geometric mean of n_1 and n_2, and make the layer 1/4 wavelength

thick (where the wavelength means that which the light has in layer number 3.)

(b) What happens at frequencies other than ν_o? (Assume that the indices of refraction are independent of frequency.) Take n_1 and n_2 to be close to one another, and neglect multiple reflections, i.e., just superpose the two most important waves. They give zero at $\nu = \nu_o$, but what do they give elsewhere? Answer:

$$R_E = \frac{2(n_1-n_2)}{n_1+n_2} \sin^2 \left(\frac{\nu-\nu_o}{\nu_o}\right). \qquad \text{(P15)}$$

Thus for incidence from air to glass without a non-reflective coating, we have $n_1 = 1$ and $n_2 = 1.5$, so that R_E is about -0.20. With a non-reflective coating designed to work perfectly at ν_o and with index of refraction independent of ν, we have

$$R_E = -0.40 \sin^2[(\nu-\nu_o)/\nu_o] = -0.4 \sin^2[(\lambda-\lambda_o)/\lambda] \quad . \qquad \text{(P16)}$$

Thus if λ_o is 5500 A^o (one A^o is 10^{-8} cm), corresponding to green, then for 4500 A^o (blue) and 6500 A^o (red) we get $R_E = -0.020$ (blue), -0.013 (red). Thus the reflected fractional intensity is 4×10^{-4} at blue, zero at green, and 1.7×10^{-4} at red.

(c) Explain the physical origin of the factor of 2 that appears in Eq. (P15).

18. Calculate <u>crudely</u> the rms time-averaged value of the electric field in space at a distance of 1 meter from a 100 watt light bulb. Don't worry about

the frequency distribution. Answer: $\langle E^2 \rangle^{1/2} \approx 2 \times 10^{-3}$ esu (statvolt/cm) ≈ 0.6 volts/cm.

*19. Show that for harmonic waves in various media the reflection coefficients for the quantities listed below all have the form

$$R\left(\begin{matrix}\text{pressure-like} \\ \text{wave}\end{matrix}\right) = \frac{(k_1/k_2) - (m_1/m_2)}{(k_1/k_2) + (m_1/m_2)} \quad , \qquad \text{(P17)}$$

where m is a "mass-density-like" quantity (aisde from proportionality factors which do not affect the ratio m_1/m_2) given in each case as follows:

medium	wave	quantity "m"	
string	transverse tension	ρ	(P18)
gas	longitudinal pressure	ρ_o	(P19)
parallel-plate transmission line	Voltage or E_x	$\mu g_x/w_y$	(P20)
coaxial transmission line	Voltage or E_r	$\mu \ln(b/a)$	(P21)
free plane wave	E	μ	(P22)

In each case the other wave quantity, the "velocity-like" wave, always has R opposite to the above.

*20. Show that if the sinusoidal spatial oscillation of the standing wave to the left of z = 0 in Fig. 23 were extrapolated to positive z, it would

have a zero at a point z_o given by

$$k_1 z_o = 1/2 \sin^{-1}\left[\frac{2k_1|k_2|}{|k_2|^2 + k_1^2}\right] . \qquad (P23)$$

Also show that, in the limit $|k_2| \gg k_1$, z_o is the mean penetration distance $|k_2|^{-1}$.

21. A "half-silvered" mirror is a piece of glass onto which a thin layer of silver (or aluminum) has been evaporated. The thickness of the layer is determined so that the energy flux emerging from the back is half of the incident flux. This is an example of barrier penetration. (If the silvered surface is thick enough, one gets total reflection.) As simplifying approximation, forget about the glass, and assume that you have a thin layer of silver of thickness s somehow mechanically supporting itself in empty space (like a piece of silver foil). (a) How thick should the foil be to transmit half of the energy flux for green light? Use the numerical quantities given in the discussion of metallic reflection of visible light from silver. (b) How thick should the foil be to transmit only 1% of the incident flux (and thus to reflect 99% of it)? (c) Suppose you wish to plate a piece of glass of dimensions 5 cm × 5 cm with a layer of the thickness s found in part (a). The glass is mounted in an evacuated enclosure at a distance of 50 cm from a tiny piece of solid silver. The silver is heated until it evaporates, spraying silver atoms uniformly in all directions. What are the dimensions of the piece of silver that is evaporated, taking it to be a cube?

* 22. Metallic conduction of solid silver, from direct current (purely resistive) through visible light (purely reactive) through ultraviolet (dispersive). (We shall ask you to derive the various steps.) We consider N electrons per unit volume, each of charge $q = -e$. They are driven by a local electric field E_x which is ultimately due to some source of harmonic waves of angular frequency ω. The field E_x is the result of both the driving radiation and the induced polarization P_x, which equals $Nqx(t)$. The dielectric constant ϵ is equal to $1 + 4\pi (P_x/E_x)$. We need $x(t)$ to find $\epsilon(\omega)$. Consider an individual electron. It is driven by the total field E_x. It is also (we assume) subject to a "drag" force proportional to its velocity, which is due to loss of energy to the ions through collisions. We define the positive constant Γ by saying that the drag force on one electron is $-\Gamma\, dx/dt$. Then Newton's second law gives

$$m \frac{d^2x}{dt^2} = qE_x - \Gamma \frac{dx}{dt} \quad . \tag{P24}$$

(a) Show that for a constant electric field, the electrons reach a terminal velocity v_o which is equal to qE_x/Γ. (b) Show that if the (previously steady) field is turned off, the electrons do not immediately come to rest, but that the velocity decays exponentially with time, with a mean decay time τ of magnitude $(\Gamma/m)^{-1}$. (c) Show that the electrical conductivity σ for a steady electric field is equal to $N e^2/\Gamma$. (The conductivity σ is defined by $J_x = \sigma E_x$. See Vol. 2, Sec. 4.3. You will learn very much of relevance to

the present problem if you will read Vol. 2, Secs. 4.3 through 4.7) Therefore if we are given the empirical value of σ and use our crude model of one conduction electron per silver atom (which gives $N = 6 \times 10^{22}/cm^3$), we can find Γ. Thus (d) plug in numbers, taking σ from a handbook of physics and chemistry. [It is indexed under resistivity and given in units of resistivity, i.e., ohm-cm. You must convert to esu, using the conversion factor 1 ohm $= 1.14 \times 10^{-12}$ sec/cm. Then recall that σ is $(\text{resistivity})^{-1}$.] At 20° C you should get $\sigma = 5.5 \times 10^{17}$ esu, and $\tau \approx 3.6 \times 10^{-14}$ sec.

After all this you can write Eq. (P24) in the equivalent form

$$m \frac{d^2x}{dt^2} = q E_x - \frac{m}{\tau} \frac{dx}{dt} \quad . \tag{P25}$$

For frequencies slow compared with $\nu_o \equiv (\tau)^{-1} = 2.8 \times 10^{13}$ cps, the electrons always rapidly reach terminal velocity during each cycle, and we have a purely resistive medium. In that limit the acceleration d^2x/dt^2 in Eq. (P25) is negligible all of the time. Of course the terminal velocity is proportional to E_x, and thus the electrons undergo harmonic motion. But their acceleration $-\omega^2 x$ is always tiny compared with the term $(\tau)^{-1} dx/dt$ in Eq. (P25) if ω is small compared with $\nu_o \equiv \tau^{-1}$. The "transition" frequency $\nu_o \approx 2.8 \times 10^{13}$ cps corresponds to a wavelength in vacuum of about 10^{-3} cm = 100,000 A° = 10 micron (all these units are commonly used.) This is in a region of wavelengths sometimes called the "far infrared" region. Thus, for any electromagnetic radiation of wavelength much longer than 10 micron, solid silver is purely resistive. For microwaves, with wavelengths

of tens of centimeters, solid silver is certainly purely resistive.

Next consider the frequency region where ω is large compared with ν_o. Then to a good approximation we can neglect the resistive damping. In that case Eq. (P25) is identical with the equation we considered for the earth's ionosphere. We then find [Chap 1, Eq. (107)] $\epsilon = 1 - (\omega_p^2/\omega^2)$. For solid silver, ν_p is 2.2×10^{15} cps, corresponding to vacuum wavelength of about 1.4×10^{-5} cm = 1400 Angstrom. This is in the "ultraviolet" region. Thus from the far infrared through the visible to the ultraviolet, silver is strongly reactive. For wavelengths shorter than 1400 A^o, i.e., frequencies above about 2.2×10^{15} cps, solid silver is transparent, i.e., it is dispersive. That is because k^2 is again positive, i.e. ϵ is real.

(e) Show that for the general case given by Eq. (P25) the dispersion relation for solid silver is given by (using the complex time dependence $e^{-i\omega t}$)

$$\frac{c^2k^2}{\omega^2} = n^2 = \epsilon = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\,\nu_o} = 1 - \frac{\omega_p^2}{\omega^2 + \nu_o^2} + \frac{i\,\nu_o\omega_p^2}{\omega(\omega^2 + \nu_o^2)} \quad , \qquad \text{(P26)}$$

where for our model ω_p is 2π times 2.2×10^{15} cps, i.e. ω_p is 1.4×10^{16} rad/sec, and ν_o is 2.8×10^{13} cps. Notice that for $\nu_o = 0$ (i.e., $\Gamma = o$) Eq. (P26) reduces to the dispersion relation previously obtained for the ionosphere. (f) Show that for microwaves Eq. (P26) can be written to a good approximation as

$$\epsilon = i\,\frac{\omega_p^2}{\nu_o\omega} \qquad \text{(P27)}$$

and k can be written [using the fact that $i^{1/2}$ is $(1 + i)/\sqrt{2}$]

$$k = \left(\frac{\omega}{\nu_o}\right)^{1/2} \frac{\omega_p}{c} \left(\frac{1+i}{\sqrt{2}}\right) \approx \left(\frac{\nu}{40\ \text{cps}}\right)^{1/2} (1 + i) \quad . \qquad \text{(P28)}$$

The inverse of the imaginary part of k is called the skin depth, δ. It is the mean penetration depth of the microwaves into the solid silver, because the wave transmitted into the reactive medium has the form

$$E_x(z,t) = B \exp(-i\omega t) \exp(ikz) \quad .$$

If we write

$$k = (\mathrm{Re}k) + i(\mathrm{Im}k) \quad ,$$

we see that we have

$$\exp(ikz) = \exp[(i\mathrm{Re}k)z] \exp[-(\mathrm{Im}k)z] \quad .$$

The skin depth δ corresponds to an attenuation factor $e^{-z/\delta}$. Thus, according to Eq. (P28), we have for silver

$$\delta = \text{skin depth in cm} \approx \left(\frac{40\,\text{cps}}{\nu}\right)^{1/2} \qquad \text{(P29)}$$

for frequencies that are small compared with $\nu_o \approx 2.8 \times 10^{13}$ cps. For 60 cps Eq. (P29) gives a skin depth of about 0.8 cm. At 100 Mc it gives a skin depth of about 6×10^{-4} cm. Thus at high frequencies the fields (and currents) are confined to a shallow "skin" near the surface of the conductor.

23. Barrier penetration by photons. [D.D. Coon, Am. Jour. Phys. 34, 240 (1966).]

Read this article. The experimental technique is ingenious and the experiment is beautiful. In the article, Coon gives an expression for the

transmitted flux which is much more complicated than ours. The reason his (exact) formula is so complicated is that when the barrier is still thin, the exponentially attenuated wave is still large when it gets to the second surface. Then it reflects, goes back to the first surface, reflects again, etc. These multiple reflections for thin barriers bring in a dependence of the fields on $e^{+|k_2|z}$, which complicates the formula. In Coon's experiment he never gets the two prisms close enough to observe this deviation from the pure $e^{-2|k_2|z}$ dependence of the energy density that one gets when the barrier is thick. (His narrowest air gap is 3.5 wavelengths of the Hg green line of 5461 A°.) (a) Use Coon's experimental Fig. 4 and our Eq. (215) plus the $e^{-2|k_{2z}|z}$ dependence for the exponential attenuation of energy density to calculate the index of refraction of the glass that Coon used in the prisms. We know that glass has $n \approx 1.5$ for visible light, so that if your answer is not near 1.5 we know that something or someone is wrong.

24. Home Exp. Rectangular two-dimensional standing surface waves on water. Obtain an icebox dish of rectangular shape (not the stiff lucite kind but the more pliable polyethylene kind). Fill it to the brim with water, and then overfill it so that the water surface bulges over the top. (This reduces the damping by the sides.) Tap gently and see the grid of standing waves. Obtain a toy gyroscope (at any toy store, or, e.g., see the catalogue of Edmund Scientific Co., Barrington, New Jersey). Hold the spinning gyroscope against the side of the dish (or, for example, couple through an inverted pie pan on which both are set). You can watch the wave-

length of the standing waves gradually lengthen as the gyro slows down.

25. Home Exp. Effective length of open-ended tube for standing waves: Use a cardboard tube from a roll of paper towels or wax paper (or a mailing tube). Use a C523.3 tuning fork as a pitch standard. Tap the open-ended tube against your head and listen. Cut off some of the tube (if necessary) so that the pitch is slightly sharp (higher pitch than 523.3). Then insert a slightly smaller tube in the end to act as a tuning element "trombone". (Use, for example, the cardboard tube from a roll of toilet paper. Cut along the length of the tube and then cut off some material so as to make a smaller tube. Then seal the crack along the seam with tape so that the smaller inner tube has no air outlets along its sides.) The mode you hear is the lowest mode of an open-ended tube. The tube contains one half wavelength of the oscillation. Sound velocity is 332 meters/sec. Therefore you "expect" the tube length to be $\lambda = 1/2\ (v/\nu) = 1/2\ (3.32 \times 10^4/523.3) = 31.7$ cm. Instead you will find that the actual length L_o is less than 31.7 cm by about 0.6 of a tube diameter. This can be interpreted as an end effect of about 0.6 of a tube radius at each end. To check that this is an end effect and not just a wrong value for the velocity of sound, try fat tubes and skinny tubes.

26. Home Exp. Resonance in cardboard tubes: Take the tube from experiment 25 (above). Hold the vibrating tuning fork near one end of the tube. If the tube's "trombone" extension is tuned for 523.3 Hz (cps), you will hear a nice loud sound. If not, vary the extension and tune for a resonance. Question: when the tube's natural vibration pitch differs from that of the tuning fork,

what pitch do you hear when you drive the tube with the fork? (First guess the answer from your knowledge of driven oscillators. Then try it.)

Here is a way to get a nice sharp resonance. Hold the tube vertically and dip the lower end into a sufficiently deep container of water. Hold the vibrating tuning fork at the open top end. Raise and lower the tube and fork to tune for resonance. The tube plus one end correction should be $\lambda/4$.

Here is another nice resonance. Fill a pop bottle about 2/3 full so that when you blow across it you get something a little higher than C523.3. Put a straw in the bottle. Hold the vibrating fork at the mouth of the bottle. Tune for resonance by sucking water out with the straw.

You can search for resonances in mailing tubes, jars, rooms,and tunnels by singing a slow ascending "siren". You will both hear and "feel" the strong resonances. The change of impedance may actually turn you off or cause you to slide to a neighboring pitch.

27. Home Exp. Beats from two tuning forks. Use two forks, each nominally C523.3. Strike one fork against the other at equal distances from the ends of the prongs. Hold both forks near one ear and listen. Repeat with the base of both forks pressed against a table (you are using the table for a sounding board). Load one prong of one fork by wrapping it with a rubber band. Vary the beat frequency by pushing the mass (the band) nearer or farther from the end of the prong. Hold both forks at one end of a cardboard tube that has been tuned to C523.3. Hold one fork at each end.

28. Home Exp. Is your sound detecting system (eardrum, nerves, brain) a phase-sensitive detector? Let us find out! Some people have said that for high frequency sound you detect the direction by noticing a time delay between a crest at one ear and a crest at the other, i.e., you detect a phase shift for the vibration of one ear drum relative to that of the other. The question of whether this is true boils down to this: Can you tell the difference between "both drums in, both drums out, both drums in, both drums out, ..." and "left drum in while right drum out, left drum out while right drum in, ... etc.".

First let us consider an open-ended mailing tube (tuned to C523 to get nice loud sounds). The tube is $\lambda/2$ in length. That means that when air is rushing to the right at the right-hand end, it is simultaneously rushing to the left at the left-hand end; i.e., the velocities at the two ends are 180° out of phase when the tube is resonating. To put it differently, the air is rushing out of both ends at the same time and into both ends at the same time. Now strike two C523 tuning forks on each other at equal distances from the ends of the prongs and put one fork at each end of the tube, so as to get beats. At a maximum in intensity each fork is helping the air go where it wants to go to resonate; i.e., at the instant when the fork at one end is pushing air into the tube, the fork at the other end is also pushing air into the tube. Half of a fast cycle (at 523 Hz) later, each fork is sucking air out of its end of the tube. Half of a beat cycle later, we have a minimum in intensity of the sound coming out of the tube. (The minimum is zero if you struck the forks at equal distances from the ends.) That is because while one fork is pushing air in at one end, the other is sucking it

out at the other, which is the opposite of what is desired to sustain the resonance. That destroys the resonance. To put it differently, motion induced by the two forks is two resonance vibrations superposed 180° out of phase, giving zero.

The point of all this is that the tube can tell whether the forks are vibrating "both in, both out ..." or instead "one in, the other out; one out, the other in; ... etc." In one case you are at a beat maximum, in the other at a minimum. So it boils down to this: if you hold one fork at one ear and the other at the other ear, will you hear beats? Or anything having the mathematical structure of beats? For example, perhaps your system will tell you, "It's coming from the left side of the room; it's coming from the right side of the room; etc.," as you go through what would correspond to beat maxima and minima. That is, if the people who say that sound direction is determined by phase differences are right, the brain might say that if one drum is a little ahead of the other in phase, say 90°, then the sound is coming from the direction of the ear drum that leads by 90°. This direction would reverse at the beat frequency. To answer the question, do the experiment.

Another way to phrase the question (with the cardboard tube as an example) is this: Do you have a hole in your head?

29. Home Exp. Measuring the relative phase at the two ends of an open tube. Suppose someone has taken a long hose-like tube, coiled it up in a box, and let one open end stick out one side of the box and the other out the other.

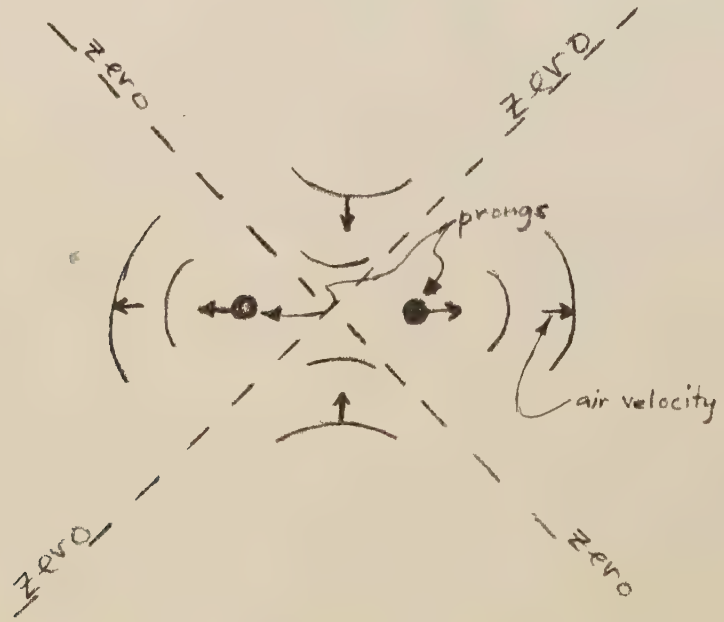
You are not allowed to see how much of the tube is coiled inside the box. By adding a small tuning trombone to a protruding end you find that you get a resonance at 523.3 cps from your tuning fork. That means that the total length is either $1/2\ \lambda$, or λ, or $(3/2)\lambda$, or ... . How can you find out whether the tube is an odd or even number of half-wavelengths? Hold two vibrating forks at one end of the tube and listen to the beats. Get the rhythm in your head so that if you remove one fork momentarily and then replace it (without disturbing the continued vibrations of both forks), you can tell that the beat maximum comes "on the beat" (in musical jargon) just where it should be. Practice several times so that you can skip a beat, count beats in your head, and come back in step when you replace the fork. (You can adjust the rubber-band loading to get a convenient beat frequency. If you find all this difficult, you can use a metronome.) Now! This time, instead of replacing the (momentarily) removed fork at the same end of the tube, carry it to the other end. Again listen for the beats. (Both forks have continued vibrating all this time.) Do they come back "on the beat", or do they come back "on the off-beat"? Depending on the experimental result, you should be able to decide whether the tube is an odd or even number of half-wavelengths. Predict the answer, then try the experiment with your half-wavelength tube. (Make another tube one wavelength long to get the opposite result.)

30. Home Exp. Radiation pattern of tuning fork. Quadrupole radiation. Hold a single vibrating tuning fork near your ear. Twirl it about its long axis

(the axis of the handle) and listen to the maxima and minima of intensity. Hold the fork at one end of a tube that has been tuned for resonance with the fork. Slowly rotate the fork about its long axis. In 360 degrees of rotation you should find four angles of zero intensity and four angles of maximum intensity. Hold the fork so that you are at a zero of intensity. Without disturbing the relative position of the fork and tube, insert a piece of cardboard that closes half of the end of the tube, giving one prong of the fork an open path down the tube but shielding the other. What happens? Why? Now strike the two forks against one another, hold one at each end of the tube, and listen for the beats. When you have established the rhythm, twist one fork by 90° about its long axis, so that you go from one angle of maximum intensity to the next. The beats do not continue on the beat. Just after the twisting, there are two beats when there should have been one, and then the beats continue on the off-beats of the original rhythm. Unless you have a good ear for keeping a steady beat, you may have trouble proving to yourself that the new beats are on the off-beats of the old, but you should have no difficulty in hearing the "extra" beat that comes when you rotate the fork. (It helps if you count "1 and 2 and 3 and ...," with the numbers on beat maxima and the "ands" on zeroes; then don't let your counting be influenced by what you hear when you twist the fork.)

What is the explanation? Think of how the forks act on the air in their vicinity. When the prongs are separating, they are pushing air away from the

outside of the prongs, giving an outwards velocity to the pushed air. Meanwhile the region between the prongs develops a slight deficit of air because the expanding prongs leave more space between them. Air rushes in from the sides to fill the deficit. The induced air velocity is thus outwards for the air lying in the plane of the two prongs and inwards for the air lying in the plane that passes between the two prongs. On the next half cycle, the prongs are coming together; the air is sucked inwards in the plane of the prongs and squeezed out into the plane between the prongs. Somewhere in between these directions there must be a direction where the induced velocity is zero, i.e., where the velocity pattern has a node. That is the explanation for the four maxima and four minima when you rotate the fork. The radiation pattern is shown in the following figure at an instant when the prongs are moving outwards:



A radiation pattern such as this is called a quadrupole radiation pattern. If you had a single prong instead of two, the radiation pattern of maxima and minima and the relative phases would be those of a dipole radiation pattern. If you take one oscillating dipole (a sound wave dipole here, but the idea applies to radio waves or any other kind of waves) and superpose its radiation on that of an identical dipole which is slightly displaced in position and oscillating 180° out of phase relative to the first, you get quadrupole radiation. Depending on the direction of the displacement relative to the pattern of waves from a single dipole, you can get many different quadrupole patterns. But they all have these common features: There are four "lobes" of strong intensity, where one gets constructive interference between the contributions from the two dipoles. The phases in neighboring lobes differ by 180 degrees. There are nodes between lobes. (A dipole pattern has just two lobes and two nodes.) The figure below shows one dipole pattern and two quadrupole patterns. Pattern (a) is a polar plot of the wave function for dipole radiation at a given instant. Pattern (b) is obtained by superposing two dipoles which are shifted slightly relative to one another along the direction of the dipole lobes and which oscillate 180° out of phase. This is the pattern of the tuning fork. Pattern (c) is obtained by superposing two dipoles which are shifted slightly relative to one another along the direction of the dipole nodes and which oscillate 180° out of phase.

(a) dipole (b) quadrupole (c) quadrupole

31. Home Exp. Overtones in tuning fork. Does your C523.3 tuning fork emit nothing but a 523 cps sound? Strike the fork against something hard. You should hear a faint high tone in addition to the strong 523 cps tone. The high tone dies away in two or three seconds. It is a higher mode of the fork and is strongly damped because it involves greater bending of the prongs. What about the note an octave higher, C1046? This is difficult to listen for because of the presence of the fundamental, 523. To search for it, use a resonating tube. Tune a tube to C1046 by tapping it on your head and listening for the octave above C523. (Or simply cut it by "theory", subtracting 0.6 of a radius R for each end from $\lambda/2$ to get the length.) Hold the C523 fork at the end of the C1046 tube and listen. (Use a tube tuned at C523 as a control. Move the fork back and forth between the C523 and C1046 tubes.)

32. Home Exp. Non-linearities in your ear. For this experiment you need one A440 and one C523 tuning fork. Bring one fork near one ear; then the other; then both. With both present you hear a pleasant interval of a minor third. What else do you hear besides A and C? Listen lower than the A. After some trying you can hear the F below A440. Altogether one has a pleasant F major triad, i.e., F, A, and C. To prove that the phenomenon happens in your eardrum and not in your brain (i.e., to prove that it is not merely that your brain just likes to hear major triads and makes up the missing part, the F), put one fork at one ear and the other fork at the other ear. (This also helps to convince you that you really hear the F.) If the phenomenon were

merely that the brain liked to fill in the chord, it could still do so. Does it? Now, let us see whether it makes sense to hear F. The difference frequency corresponding to A440 and C523 is 523-440 = 83 cps. This is a low E, two E's below middle C. (Pitches of musical notes can be found in a handbook of physics and chemistry indexed under musical scales.) But the F that I hear is the F below A440. It is F349 (according to my handbook). Where does that come from? It's very simple: two times 440 minus 523 is 357. We are hearing a non-linear difference tone (see Chap. 1, Sec. 1.5) between the octave overtone of A440 and the fundamental of C523. At least that is what I hear. (Why don't I hear E83? I don't know.) Note that 357 is slightly above F349. The next note above F349 is F sharp 370. Thus 357 is a fraction 8/21 of the way between F and F#. It is a slightly sharp F. (It also sounds a little sharp to me. I am convinced!)

Presumably the way to get easily audible difference tones without resorting to overtones of the forks is to choose two forks whose difference is in an easy range to listen to. Perhaps by using a table of pitches and a catalogue (e.g., Edmunds) of available tuning forks you can find a good combination.*

33. Home Exp. Phase of beats between two forks. Strike two C523 forks together and listen to the beats, using a resonating tube as an amplifier for ease of listening. Once you have the rhythm so you can count it out loud (without the forks vibrating), count several times before striking the forks, then strike them on each other on the beat. You will find that the beat

---

* Added note: I tried it. It doesn't work! My revised explanation for the "filling in of the tonic note of a major triad" is that it is due to a cubic non-linearity in my eardrum. You may easily show that $(\cos \omega_1 t + \cos \omega_2 t)^3$ contains the term $\cos(2\omega_1 - \omega_2)t$, among others. The quadratic non-linearity seems to be absent. I can hear $2\omega_1 - \omega_2$, but not $\omega_1 - \omega_2$!

maxima always have the same phase relation to the striking time. Why is this? Does time zero when the forks strike one another correspond to a beat maximum or minimum? Why?

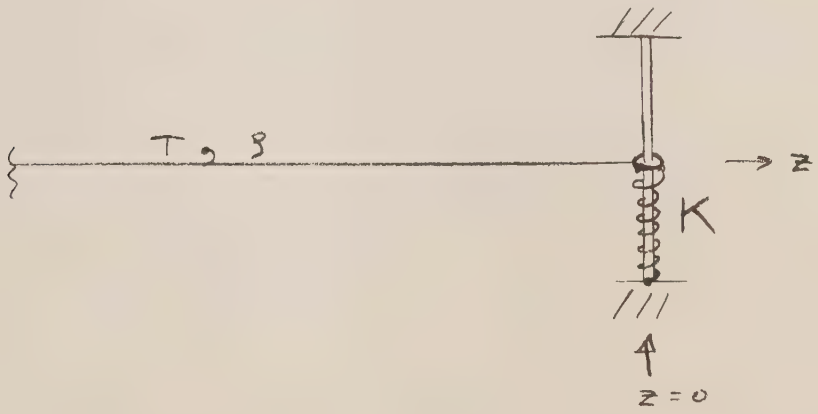
34. Home Exp. Standing waves in water.

(a) Dip a vibrating tuning fork in water and look, especially between the prongs.

(b) Lay a vibrating tuning fork flat on water (like two parallel floating logs) and look between the prongs. (Some of the modes are rapidly damped. There is one which persists for several seconds.) Try illuminating with a small light source at various angles (parallel and perpendicular to the prongs) to see the amazing amount of structure.

(c) Fill a styrofoam cup to the brim and over, so the water bulges above the top. Tap the cup gently and watch. (Styrofoam cups are the best. They are more elastic than paper, so they do not damp the waves as rapidly. They are not massive, so that it is easy to couple the driving forces from light objects like pencils.) It also works well when not overfilled. (Overfilling reduces the damping.) A styrofoam ice-bucket is a nice large system. Fill it over the brim and drive it with a spinning gyroscope. To make the gyroscope vibrate, unbalance its rotor by loading one part of the rim with some tape. To drive an ice-bucket full of water it helps to have the larger size toy gyro (about one dollar in a toy store -- or see, e.g., Edmund's catalog). As the gyro slows down you can watch the wavelength grow, and the motion should also pass through at least one resonance where the wave amplitude becomes large.

35. Consider a string with tension T and density ρ stretched from $z = -\infty$ to $z = 0$, where the string is attached to a massless ring that slides on a frictionless transverse rod. (So far, we have a "free" end of a string.) Now add a massless spring of spring constant K to provide a transverse return force, as shown below.



The spring absorbs no energy permanently (although it does whenever it is temporarily compressed or stretched). Thus for a harmonic wave the reflected energy flux must equal the incident flux when we average over a complete period. At any instant, energy may be flowing into the spring as it is stretched (or compressed) from its equilibrium position. This fact introduces a time delay -- the reflected wave has the same amplitude but not the same phase constant. Therefore we can write (using complex wave functions, of which the real parts give the actual physical quantities)

$$\psi_{inc} = Ae^{-i\omega t}\, e^{ikz} \quad , \qquad \text{(P30)}$$

$$\psi_{ref} = R_\psi A e^{-i\omega t} e^{-ikz} \quad . \qquad (P31)$$

(a) Show that the boundary condition at z = 0 is given by

$$T \frac{\partial \psi(z,t)}{\partial z} + K\psi(z,t) = 0 \quad . \qquad (P32)$$

(b) When dealing with resistive terminations we found that the reflection coefficient R_ψ is equal to $(Z_1-Z_2)/(Z_1+Z_2)$. Use this formula to define the load impedance Z_2 of the terminating spring. Find R_ψ and solve for Z_2 to obtain

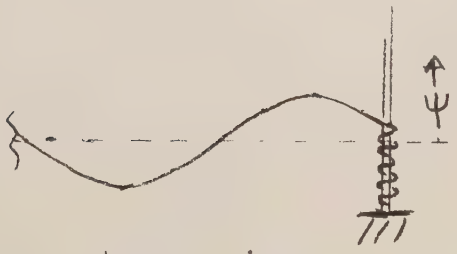
$$Z_2(\text{spring}) = \frac{1}{-i\omega K^{-1}} \quad . \qquad (P33)$$

This answer is to be compared with the load impedance offered by a capacitance,

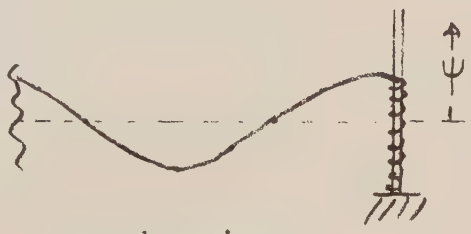
$$Z_2(\text{capacitance}) = \frac{1}{-i\omega C} \qquad (P34)$$

[See Vol. 2, table on page 289. There the time dependence $e^{+i\omega t}$ is used, whereas we used $e^{-i\omega t}$. That is why we have $-i\omega$ (going with $e^{-i\omega t}$) wherever Vol. 2 has $+i\omega$ (going with $e^{+i\omega t}$). Both conventions are commonly used. You have to know which one is being used; otherwise you will end up going backwards in time instead of forwards. We see that a stiff spring is like a small capacitance. A weak spring is like a large capacitance. In the limit of a very weak spring, the end of the string is free. Then the slope goes to zero at the end. (See the figure below.) In the limit of a very large capacitance (as the termination of a transmission line), one has a short circuit. Then the voltage and electric field go to zero. Thus we see that

the electric field behaves like the string slope. Is a weak spring (for the string) like metallic reflection for visible light or reflection from the ionosphere for radio waves? No, because in the case of the weak spring the place where the string slope goes to zero is <u>on</u> the string a slight distance before the end. (See the Figure below.) However for the case of metallic reflection the place where the electric field extrapolates to zero is slightly <u>inside</u> the metal. That is, the sign of the small phase shift that gives the deviation of R from R = -1 is opposite for the two cases. If you want to find the string (and transmission line) analogue for metallic reflection, you need something else. (See problem 36.)

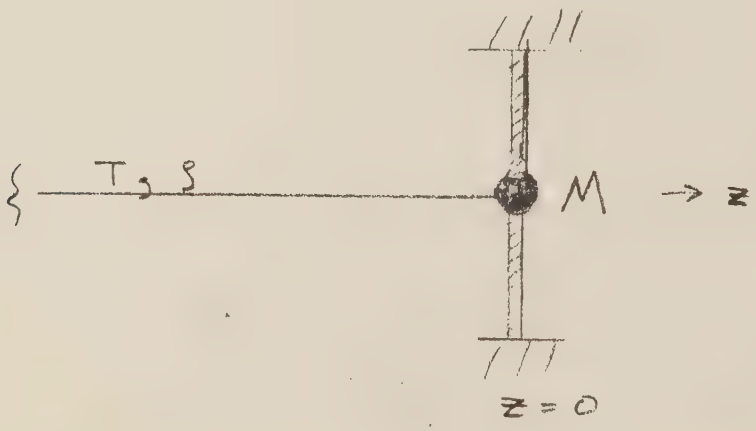


strong spring



weak spring

36. Consider a string stretched from $z = -\infty$ to $z = 0$, where it is attached to a mass M. The mass M has a hole through which passes a frictionless transverse rod on which M can slide, as shown below.



The mass M absorbs energy (which it carries as kinetic energy) and also delivers energy back to the string when it slows down. We expect that we have a purely reactive load, just as for the spring (problem 35), so that the reflected wave has the same amplitude as the incident wave and a phase shift relative to the incident wave.

(a) Show that the boundary condition at $z = 0$ is given by

$$T \frac{\partial \psi(z,t)}{\partial z} + M \frac{\partial^2 \psi(z,t)}{\partial t^2} = 0 \quad . \tag{P35}$$

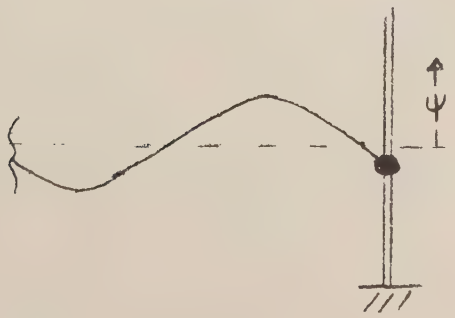
(b) Show that if we write ψ_{inc} and ψ_{ref} in the form of Eqs. (P30) and (P31) and write $R_\psi = (Z_1 - Z_2)/(Z_1 + Z_2)$, with Z_1 the characteristic impedance of the string, T/v, then the impedance Z_2 of the mass M is given by

$$Z_2(\text{mass}) = -i\omega M \quad . \tag{P36}$$

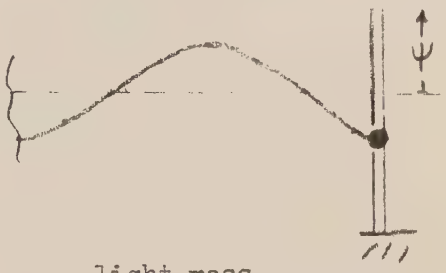
This answer can be compared with the load impedance offered by an inductance L,

$$Z_2(\text{inductance}) = -i\omega L \quad .$$

(Vol. 2, table on page 289 gives Z = + iωL because the convention $e^{+i\omega t}$ is used for the complex time dependence. We use $e^{-i\omega t}$ and thus get -iω wherever +iω appears in that table.) Thus a big mass is like a big inductance. An infinite mass holds the string velocity at zero at z = 0, just as an infinite inductance as the termination of a transmission line holds the current at zero. An infinite mass reflects $\partial\psi/\partial z$ with coefficient +1, just as an infinite inductance reflects the voltage. A small mass allows large velocity (small inductance allows large current). A small mass holds the string slope near zero, with the nearest zero of slope slightly to the right of z = 0, "inside the mass" (i.e., inside a virtual string to the right of z = 0). (See Figure below.)



heavy mass



light mass

Similarly for a small inductance one has almost a short circuit. The voltage and electric field are held almost at zero and hence reflect with a minus sign plus a small phase shift. The phase shift is such that the nearest zero of voltage or electric field occurs slightly "inside the inductance". That is also what one finds for reflection of radio waves from an idealized ionosphere (idealized so that it starts abruptly at $z = 0$) or for reflection of visible light from a metal. The electric field extrapolates to zero slightly inside the metal. That is, the plot of electric field looks like Fig. 23. Thus metallic reflection is "just like" reflection at an almost free end of a string terminated with a small mass, or it is like reflection at the end of a transmission line terminated by a small inductance. As the mass goes to zero, or the inductance goes to zero, or the plasma frequency goes to infinity (number of free electrons per unit volume goes to infinity), the behavior is analogous in the three cases, if we take string slope, voltage, and electric field to be analogues.

37. For metallic reflection of light with frequency far below the plasma oscillation frequency but with frequency high enough to allow us to neglect any resistive damping, show that the reflection coefficient of the electric field is given by $R_E = -(Z_1-Z_2)/(Z_1+Z_2)$, provided that we take Z_1, the characteristic impedance for vacuum, to be $Z_1 = 4\pi/c = 377$ ohms per square cm, and provided that we take Z_2, the effective impedance presented by the metal or plasma, to be

$$Z_2(\text{plasma}) = -i\omega \frac{4\pi}{c\omega_p} = -i\omega \frac{377 \text{ ohms}}{\omega_p} \quad . \qquad \text{(P37)}$$

Thus the metal (or plasma) acts like an effective inductance L_{eff}, with

$$L_{eff} = \frac{377 \text{ ohms per square cm}}{\omega_p} \quad . \qquad \text{(P38)}$$

38. Consider a string of alternating masses and springs. The masses all have mass M, and the springs all have the same spring constant C and the same unstretched length a. At equilibrium the springs are unstretched. Consider longitudinal waves of displacement of the masses along the direction of the string. Take the limit where neighboring masses have nearly the same motion so that you can approximate the displacement of the masses in the neighborhood of z at time t by the continuous function $\psi(z,t)$. Derive the classical wave equation for $\psi(z,t)$, and show that the wave velocity is given by

$$v^2 = \frac{Ca^2}{M} \quad . \qquad \text{(P39)}$$

39. Sketch the behavior of the velocity and of the slope of a pulse on a string under reflection at a free end if the behavior of the displacement is as shown in Fig. 8. That is, construct diagrams analogous to those shown in Figs. 6(b) and 6 (c).

*40. Index of refraction of glass for visible light.

Let us make a simple classical model of a "glass molecule" (glass is mostly silicon dioxide). The model is this. Each glass molecule has a number v (v is for valence) of "outer shell" or "optical" electrons. (We leave v a free parameter to be determined by experiment.) Each electron acts independently. Each electron is bound to a massive nucleus by a spring having

a spring constant such that the natural oscillation angular frequency (for small oscillations of the electron) is ω_o. When the molecule is immersed in an electric field E_x, the electron experiences a force qE_x from the field (with $q = -e = -4.8 \times 10^{-10}$ esu) and a force $-m\omega_o^2 x$ from the spring.

(a) Write down Newton's law for the motion of the electron. Assume that E_x is a harmonic oscillation of angular frequency ω. Show that the steady state solution (we are neglecting all damping) is

$$x(t) = \frac{q\,E_x(t)}{m(\omega_o^2 - \omega^2)} \quad . \qquad \text{(P40)}$$

This gives polarization $P_x = Nqx$, where N is the number of "optical electrons" per unit volume in the glass. This gives dielectric constant $\epsilon = 1 + 4\pi(P_x/E_x)$, and thus gives index of refraction $n^2 = \epsilon$, i.e., show that

$$n^2 = 1 + \frac{4\pi\, Ne^2/m}{(\omega_o^2 - \omega^2)} = 1 + \frac{\omega_p^2}{(\omega_o^2 - \omega^2)} \quad , \qquad \text{(P41)}$$

where we give the name ω_p^2 to $4\pi\, Ne^2/m$, even though it is certainly not a "plasma frequency". Now we wish to find ω_p^2 and ω_o^2 for glass and visible light. We can take values of n from a handbook of physics and chemistry. We need two equations in two unknowns. One is Eq. (P41). The other is obtained by considering the variation of the index of refraction with wavelength, i.e., the dispersion. (b) Show that ω_p^2 and ω_o^2 are given by

$$\omega_p^2 = (n^2-1)^2 \frac{\omega^2}{n} \left[\frac{-d\lambda}{\lambda dn}\right] \quad , \qquad \text{(P42)}$$

$$\omega_o^2 = \omega^2 + \frac{\omega_p^2}{n^2 - 1} \qquad \text{(P43)}$$

(c) Now take the following experimental values for "zinc crown glass" (the wavelengths are given in microns; one micron is 10^{-4} cm).

Table P1

Color	Source	λ (microns)	n observed	n predicted
near ultraviolet		0.361	1.539	1.540
dark blue	H_γ	0.434	1.528	1.529
blue-green	H_β	0.486	1.523	(1.523)
yellow	sodium	0.589	1.517	(1.517)
red	H_α	0.656	1.514	1.515
very dark red		0.768	1.511	1.513
infrared		1.20	1.505	1.507
infrared		2.00	1.497	1.505

The three sources H_α, H_β, and H_γ are the first three spectral lines of the Balmer series for hydrogen. Their wavelengths satisfy Bohr's famous formula

$$\lambda^{-1} = 109{,}678 \left[\frac{1}{2^2} - \frac{1}{N^2}\right] \quad , \quad N \text{ integral} \quad , \qquad \text{(P44)}$$

with λ^{-1} in inverse centimeters and with N = 3 (for H_α), 4 (for H_β), 5 (for H_γ), ... etc. The constant 109,678 cm^{-1} is called Rydberg's constant. The term $1/2^2$ indicates that the Balmer series consists of transitions from excited states N = 3, 4, 5, ... to the first excited state above the ground state

of the hydrogen atom. (The ground state of hydrogen has N = 1, the first excited state has N = 2.) The yellow source is the sodium D line (actually this is a "doublet" of two closely-spaced spectral lines). This is what gives the yellow color when you throw table salt on a gas flame.

Use Table P1 to find ω_p and ω_o, using the entries for blue-green and yellow, which have an average wavelength of 0.538 micron (greenish yellow), corresponding to a frequency 5.58×10^{14} Hz (cps), i.e., $\omega = 3.50 \times 10^{15}$ rad/sec. Show that Eqs. (P42) and (P43) then give

$$\omega_p = 2.04 \times 10^{16} \text{ rad/sec}, \quad \omega_o = 1.815 \times 10^{16} \text{ rad/sec} \quad . \quad \text{(P45)}$$

Now comes the question: How good is the formula, Eq. (P41)? We have used the two middle entries in Table P1 to find the constants ω_p and ω_o. Let us now predict what n will be for other entries in the table. That is how the column "predicted" was made. Thus (d) verify one of the entries in the "predicted" column.

Now use the fact that the density of glass is about 2.6 gm/cm^3, and assume "zinc crown glass" is just SiO_2 with molecular weight 60 gm/mole. Then (e) show that the effective number of "optical electrons" per glass molecule (given our result for ω_p^2) is equal to 1.99 optical electrons per glass molecule of molecular weight 60. Since silicon has valence 4 and likes to donate two electrons to each oxygen atom, we might have expected the number of optical electrons to be 4 instead of 2. Maybe each electron is only "half-effective" or something. Let us not push this simple non-quantum-mechanical model too hard.

Chapter 4 - Polarization

Section		Page
4.1	Introduction.	1
4.2	Description of polarization.	6
4.3	Production of polarized transverse waves.	15
4.4	Double refraction.	29
4.5	Bandwidth, coherence time, and polarization.	40
	Problems and Home Experiments.	51

4.1 Introduction.

We learned in Chapter 3 that electromagnetic waves have their electric and magnetic fields transverse to the direction of propagation $\hat{z}$. There are two transverse directions, $\hat{x}$ and $\hat{y}$, and the fields with one orientation with respect to $\hat{x}$ and $\hat{y}$ are independent of those with an orientation differing by 90°. One may therefore have various amounts (amplitudes) of the fields in each of the two transverse directions and various possible relative phases. A specific relation of the amplitudes and phases of the two independent transverse fields is called a state of polarization.

When electromagnetic waves impinge on (and interact with) matter, it often happens that different polarization states of the incident radiation do not interact with the material in the same way. For example, we might find a material such that charged particles in the material are free to move along $\hat{x}$ but cannot move at all along $\hat{y}$. In that case E_x can do work on the charged particles, but E_y cannot. Then the electromagnetic wave energy associated with E_x may be reduced by being converted into charged-particle kinetic energy and thence by particle-particle collisions into heat energy, whereas the amplitude of E_y is not affected. Or instead it may only happen that the phase of E_x is shifted relative to that of E_y, without any diminution of energy (i.e., without decrease in the amplitude of E_x). In all such cases of asymmetrical interactions, the polarization state of the electromagnetic radiation is modified or changed by the interaction. This fact has many important consequences. By studying the effect of well-understood materials on an incident beam of unknown polarization state, the polarization state can be determined. Conversely, by measuring the modification of a known polarization state by a material, one can learn something

about the material. For example, the direction of magnetic field in "our" spiral arm of our galaxy is now being mapped out by measuring the polarization direction of radio waves from extra-galactic sources as a function of the direction of the source and of the wavelength of the radiation (G. L. Berge and G. A. Seielstad, Scientific American, June 1965, p. 46).

In Sec. 4.2 we examine ways of describing the polarization of transverse electromagnetic waves and transverse waves in a slinky. (A slinky is a long spring, available in toy stores for about $1.00.) The most general polarization state for monochromatic harmonic waves turns out to be a motion of the displacement in a transverse elliptical path. This is called elliptical polarization. Linear transverse polarization (i.e., motion with the displacement zero along one transverse direction) can be thought of as a motion along a degenerate ellipse. Circular polarization is also a special case of elliptical motion.

In Sec. 4.3 we examine several methods for producing a desired polarization state. It is easiest to control the polarization if one has control of the radiation process, as one does when one shakes a slinky or broadcasts electromagnetic waves from an antenna of one's own design. If one starts with light from a light bulb or the sun, the problem is one of somehow selecting a desired polarization state from the complicated superposition of different states that is present. One can absorb the undesired polarization components, as is done by a sheet of polaroid. Instead one can arrange to get negligible reflection of the undesired polarization components and then look only at the reflected radiation. That is what makes the blue sky polarized and what makes light reflected from water, or glass, or concrete, or a knee polarized.

In Sec. 4.4 we examine ways to change the relative phase of different components of polarized light, so as to change the state of polarization.

If one has control of the emission process it is easy. For example if one is emitting microwaves along $\hat{z}$ from two crossed straight antennas, one along $\hat{x}$ and one along $\hat{y}$, one can change the polarization state by changing the relative phase of the currents driving the two antennas. (For example, one may use a coaxial transmission line to carry the current from the oscillator to each antenna and vary the length of one of the transmission lines to obtain a variable time delay and hence a variable phase shift.)

Instead one may introduce relative phase shifts by passing the waves through a medium whose phase velocity depends on the direction of transverse vibration. Such a medium is said to be doubly refracting. One can make a doubly refracting medium for waves on a string or slinky by making either the inertia or the return force for x displacement be different from that for y displacement (the string is along $\hat{z}$). To make the inertia be greater for x displacement than for y displacement, one could lay a massive "outrigger" string on each side of (and parallel to) the original string and couple all three strings with massless transverse push rods. For horizontal vibrations the outrigger strings on either side are pushed and their mass must be displaced if the central string is to move. For vertical vibrations of small displacement the outrigger strings barely move. Thus the mass per unit length that must move differs for the two cases. Similarly one could add extra transverse springs in one direction, so as to make the return force depend on transverse direction. A more practical way to devise a doubly refracting medium is to use a solid rod of oblong cross section instead of a flexible string. The phase velocity of transverse waves propagating along a rod of oblong cross section is different for transverse polarizations along the short and long edges, because the stiffness in one

direction is greater than that in the other. If one drives such a rod at one end with a linearly polarized transverse motion which is not aligned along either of the edges of the rod but lies somewhere in between, the components with polarization along the two edges travel with different phase velocities and hence get out of phase by an amount that increases linearly with the distance along the rod. The motion at the other end of the rod can be arranged to be anything we please by selecting the proper orientation of the rod (with respect to the driving force) and the proper length. Thus, for example, we may convert linear transverse polarization into circular polarization.

The corresponding device in the case of electromagnetic radiation is called a retardation plate. Your optics kit contains a quarter-wave plate and a half-wave plate, which introduce relative phase shifts of $\frac{1}{2}\pi$ and π, respectively, between the two transverse linear polarization components of visible light. Such a device is made by using doubly refracting transparent materials like cellophane or Iceland spar (calcite crystals) whose index of refraction is different for different directions of displacement of the driven electrons.

Another type of double refraction is exhibited by sugar molecules in solution. These molecules are helical in shape and have a different index of refraction for right-handed circularly polarized light than for left-handed. (These differ by 180° in the choice of the relative phase of vibrations in the two independent transverse directions.) This leads to a phenomenon called optical activity.

Another process similar to double refraction is obtained by reflecting light from a shiny metallic surface. The exponential "penetration depth"

differs for the polarization components parallel and perpendicular to the plane of incidence (the plane containing the incident ray and the normal to the surface). Therefore there is a relative phase shift introduced between these two components in the reflected light. By starting with light linearly polarized at 45° to the plane of incidence and looking at the reflected light at a judiciously chosen angle of incidence, one can obtain circularly polarized light, with the handedness depending on how one tilts the first polaroid (i.e., depending on whether it is 45° "to the right" or "to the left").

In Sec. 4.5 we examine the relation between the frequency bandwidth $\Delta\nu$ of a superposition of waves and the "coherence time," which is defined as the time interval over which the relative phases of independently produced polarization components can be expected to remain roughly constant. (If the x and y polarization components are not produced independently then their relative phase need not change.) This examination enables us to say what we mean by "unpolarized light." We conclude that for monochromatic waves, i.e., zero bandwidth, there are no such things as "unpolarized waves." A monochromatic harmonic wave is always in a "100% pure" or "completely polarized" state. But if one has a band $\Delta\nu$ of frequencies (as one always does in practice, since no experiment or process goes on and on forever like a pure harmonic wave does), then after a time of approximately $(\Delta\nu)^{-1}$ the various frequency components in the band are no longer in exactly the same phase relation as they were at the beginning of the interval. Therefore the polarization can change, if the x and y polarization components are not "locked" in phase by some means. Over a long time one can then find intervals

where every kind of oscillation has occurred. If one has not followed the process through its evolution, one expresses one's ignorance by saying that the resulting waves are "unpolarized." All that one has to do to turn "unpolarized" light into "strongly polarized" light is to learn how to measure the polarization in a time short compared with $(\Delta\nu)^{-1}$, before the polarization has a chance to change.

In Sec. 4.5 we also examine an application of the idea of coherence time to help understand the phenomenon of beats. The role of the coherence time is easily understood in the case of coupled identical pendulums with damping. If one (or both) normal modes are damped out in a time comparable to or shorter than the period for beats, it will be very difficult to observe the beats. To get nice beats one should have damping times that are long compared with the beat period. The inverse of the damping time of a mode is approximately equal to the frequency bandwith $\Delta\nu$. Therefore the requirement that the damping time be long compared to the beat period is equivalent to the requirement that the frequency bandwidth of each mode be small compared with the beat frequency. The damping time is the "coherence time" $(\Delta\nu)^{-1}$.

4.2 Description of polarization

All the waves that we study consist of some physical quantity whose displacement from its equilibrium value varies with position and time. The displacement can be described by a vector $\underset{\sim}{\psi}(x,y,z,t)$. We usually study plane waves for which $\underset{\sim}{\psi}$ has the form $\underset{\sim}{\psi}(z,t)$, where z is measured along the propagation direction. (Here we include both standing waves and traveling waves.) The quantities $\partial\underset{\sim}{\psi}(z,t)/\partial t$ and $\partial\underset{\sim}{\psi}(z,t)/\partial z$ are often the quantities with the most interesting physical properties. We have seen this

to be the case for waves on a string and for sound waves, where in each case $\underset{\sim}{\psi}(z,t)$ designates the displacement of the particles of the medium away from their equilibrium positions. [In the case of electromagnetic plane waves the interesting physical quantities analogous to $\partial\underset{\sim}{\psi}/\partial z$ and $\partial\underset{\sim}{\psi}/\partial t$ are $\underset{\sim}{E}(z,t)$ and $\underset{\sim}{B}(z,t)$. The quantity that is somewhat analogous to the displacement $\underset{\sim}{\psi}(z,t)$ is the vector potential $\underset{\sim}{A}(z,t)$. The vector potential is defined so that the electric field $\underset{\sim}{E}(z,t)$ is equal to $-c^{-1}\,\partial\underset{\sim}{A}(z,t)/\partial t$ and the magnetic field $\underset{\sim}{B}(z,t)$ is equal to $\nabla \times \underset{\sim}{A}$. For plane waves $\underset{\sim}{B}$ does not depend on x and y, so that $\underset{\sim}{B}(z,t)$ is $\hat{z} \times \partial\underset{\sim}{A}/\partial z$. Note that transverse displacement implies transverse fields, and that transverse fields require only transverse displacement.]

For plane waves propagating along $\hat{z}$ we can write the displacement as

$$\underset{\sim}{\psi}(z,t) = \hat{x}\,\psi_x(z,t) + \hat{y}\,\psi_y(z,t) + \hat{z}\,\psi_z(z,t) \quad . \quad (1)$$

In the case of transverse waves on a string, $\underset{\sim}{\psi}$ has only x and y components. Such waves are said to have transverse polarization. (Actually, one can also have longitudinal waves on a string consisting of variations of the tension and of the longitudinal velocity of the particles of the string.) For sound waves, the displacement $\underset{\sim}{\psi}$ is along the propagation direction $\hat{z}$. One says that these waves are longitudinally polarized. (Actually, it is possible to have transverse sound waves in a tube. These transverse waves can be thought of as longitudinal waves that are not headed down the tube but bounce from one side of the tube to the other. The net propagation direction is down the tube, but the air oscillations have transverse components as well as longitudinal components.) In the case of electromagnetic plane waves, the displacement $\underset{\sim}{\psi}$ is transverse to $\hat{z}$, as we saw in Sec. . We found there

that $\underset{\sim}{E}$ and $\underset{\sim}{B}$ are always transverse to $\hat{z}$, for plane waves in vacuum. (It is possible to have longitudinal components of $\underset{\sim}{E}$ and $\underset{\sim}{B}$ if, for example, the the waves are enclosed in a wave guide or cavity.) In water waves the water undergoes both transverse (up and down) and longitudinal displacements.

Linear polarization. In the case of transverse waves (electromagnetic plane waves and transverse waves on a string), if the displacement is an oscillation back and forth along a fixed line transverse to $\hat{z}$, the waves are said to be linearly polarized. Such a direction can be $\hat{x}$, or $\hat{y}$, or a general fixed transverse direction $\hat{e}$ given by

$$\hat{e} = \hat{x} \cos \phi + \hat{y} \sin \phi \qquad (2)$$

where ϕ is an arbitrary (but fixed) angle. See Fig. 1.

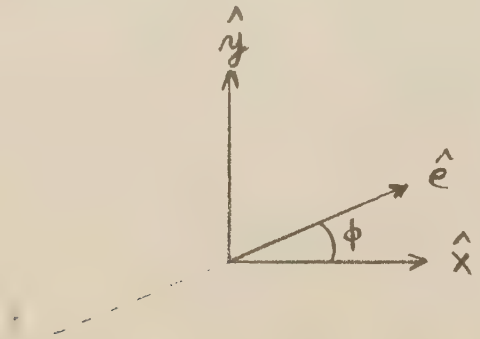


Fig. 4-1. Linear polarization.

There are two independent transverse directions. These may be taken to be $\hat{x}$ and $\hat{y}$. More generally one direction may be $\hat{e}$ as given by Eq. (2) for an arbitrary choice ϕ, with the second direction (orthogonal to the first) obtained by replacing ϕ by $\phi + 90^\circ$. At fixed z (where we need not be concerned about whether we are dealing with standing or traveling

waves or both), the oscillations corresponding to a linearly polarized plane wave can be written in the form (suppressing z from the notation)

$$\underset{\sim}{\psi}(t) = \hat{e}\, A \cos \omega t \quad , \tag{3}$$

with $\hat{e}$ given by Eq. (2). More specifically, a linearly polarized standing wave can take the form (assuming a node in $\underset{\sim}{\psi}$ at $z = 0$)

$$\underset{\sim}{\psi}(z,t) = \hat{e}\, A \sin kz \cos \omega t \quad . \tag{4}$$

A linearly polarized traveling wave (traveling in the $+z$ direction) has the form

$$\underset{\sim}{\psi}(z,t) = \hat{e}\, A \cos (\omega t - kz) \quad . \tag{5}$$

(As usual, we suppress unnecessary phase constants.)

Circular polarization. If the displacement in a transverse wave is a motion in a circle, the waves are said to be circularly polarized. We can define a rotating unit vector $\hat{e}_+$ rotating in the $+\hat{z}$ direction (the rotation direction is defined by a right-hand rule) and a rotating unit vector $\hat{e}_-$ rotating in the $-\hat{z}$ direction by

$$\hat{e}_+ = \hat{x} \cos (\omega t + \phi_+) + \hat{y} \sin (\omega t + \phi_+) \quad , \tag{6}$$

$$\hat{e}_- = \hat{x} \cos (\omega t + \phi_-) - \hat{y} \sin (\omega t + \phi_-) \quad . \tag{7}$$

(We can set ϕ_+ or ϕ_- equal to zero by properly setting the clock, if we so desire. That is what we do in Fig. 2.)

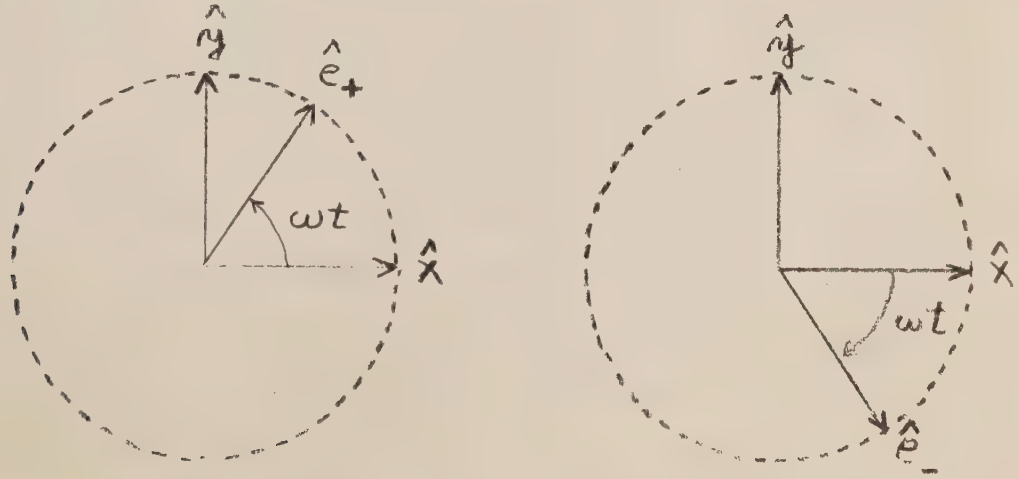


Fig. 4-2. Circular polarization

In $\hat{e}_+$ the $\hat{x}$ vibration leads the $\hat{y}$ vibration by 90° in phase, and in $\hat{e}_-$ the $\hat{x}$ vibration lags behind the $\hat{y}$ vibration by 90°. Notice that Eq. (6) can be obtained by replacing ϕ in Eq. (2) by $\phi + \omega t$. Similarly Eq. (7) can be obtained from Eq. (2) by replacing ϕ by $-\phi - \omega t$.

For fixed z the oscillation corresponding to circular polarization directed along $+\hat{z}$ has the form (suppressing z from the notation and setting ϕ equal to zero)

$$\psi(t) = \hat{e}_+ A = A [\hat{x} \cos \omega t + \hat{y} \sin \omega t] . \qquad (8)$$

A circularly polarized standing wave with circular polarization along $+\hat{z}$ (and with a node at $z = 0$) has the form

$$\psi(z,t) = \hat{e}_+ A \sin kz = A [\hat{x} \sin kz \cos \omega t + \hat{y} \sin kz \sin \omega t] . \qquad (9)$$

A circularly polarized traveling wave with circular polarization along $+\hat{z}$ (and traveling in the $+\hat{z}$ direction) has a form that can most easily be obtained by replacing ωt by $\omega t - kz$ in Eq. (8). Then we have

$$\psi(z,t) = A [\hat{x} \cos (\omega t - kz) + \hat{y} \sin (\omega t - kz)] . \qquad (10)$$

One way to have a tangible experience of transverse polarization is to play with a "slinky" (obtainable in any toy store; the 3 inch diameter ones are better than the $1\frac{1}{2}$ inch diameter ones). To obtain standing waves, tie one end to a telephone pole and shake the other. To simulate a "free" end, tie one end of the slinky to about 30 feet of string, and tie the other end of the string to the pole. Linearly or circularly polarized standing waves are easy to generate. Traveling harmonic waves are difficult to generate, because it is not easy to terminate a slinky with an impedance-matching resistive load. [We suspect that the combination of a long string (to "free"

the end) with suitable styrofoam("massless") pistons or paddles stirring buckets of water can be made to work.] However one can easily send a pulse or wave packet down the slinky and follow it as it reflects at fixed or free ends.

By playing with a slinky or by studying the above equations you can verify the following (which hold for electromagnetic plane waves as well):

1. In a linearly polarized wave the displacement at fixed z goes through zero twice per cycle. In a standing wave all points go through zero simultaneously. In a traveling wave all points undergo the same motion as all others, but with a phase shift corresponding to the propagation time of the wave between the points.

2. In a circularly polarized standing or traveling wave the displacement at fixed z has constant magnitude. If the slinky is carrying a circularly polarized traveling wave, a snapshot at a fixed time t would show the slinky to be in the shape of a corkscrew. If instead it carries a circularly polarized standing wave, the slinky lies completely in a plane. The single snapshot could not distinguish this shape from either that of a linearly polarized standing wave or that of a linearly polarized traveling wave. A second snapshot taken slightly later,combined with the first snapshot,would tell us which of these three possibilities held.

3. The reflection at a termination of a wave packet with circular polarization along $+\hat{z}$ gives a reflected wave with circular polarization along $+\hat{z}$. This is true for reflection from either fixed or free ends (or any other kind of load). Thus the sense of the rotation with respect to the z axis is preserved under reflection. This follows directly from conservation of angular

momentum (because you exerted a torque when you generated them, and they will exert a torque if they are absorbed). The same thing holds for circularly polarized electromagnetic waves: they carry angular momentum about the z (propagation) axis. (See Advanced Topic 8.)

4. A traveling corkscrew wave packet that is going away from you on a slinky and rotating clockwise as seen from your vantage point has at any instant the form of a left-handed screw. This kind of circularly polarized traveling wave is therefore called a left-handed circularly polarized traveling wave. Suppose it now reflects from the far end of the slinky and comes back towards you. The sense of rotation is not changed; it still appears to rotate clockwise as it comes toward you. But now it has the form of a right-handed screw. It is said to be right-handed circularly polarized. (If you think of a corkscrew shape undergoing reflection, the handedness is reversed because the corkscrew is "turned inside out" by the reflection.) Electromagnetic radiation has the same behavior as the slinky. By this we mean that the sense of rotation with respect to $\hat{z}$ of circularly polarized light or microwaves or any other electromagnetic radiation is not changed by a reflection, but the handedness, i.e., the sense of rotation with respect to the propagation direction, is reversed. The fact that the handedness of light is reversed under reflection is not new to you. You know that if you look at your right hand in a mirror it looks like a left hand. That may not seem to be obviously related to the preservation (under reflection from a mirror) of the sense of rotation with respect to $\hat{z}$ of circularly polarized light, but in fact it is related. (Both can be thought of as being the result of conservation of momentum transverse to $\hat{z}$ in the interaction of the wave with the reflecting medium, i.e., the wall in the case of the slinky or the electrons of the mirror in the case of electromagnetic radiation.)

General transverse polarization. Elliptical polarization. At a fixed z a general oscillation with transverse polarization has the form

$$\underset{\sim}{\psi}(t) = \hat{x}\, A_1 \cos(\omega t + \phi_1) + \hat{y}\, A_2 \cos(\omega t + \phi_2) \quad . \quad (11)$$

If ϕ_2 equals either ϕ_1 or $\phi_1 \pm \pi$, we have linear polarization. If ϕ_2 is $\phi_1 - \frac{1}{2}\pi$ and A_2 equals A_1, we have circular polarization along $+\hat{z}$. If ϕ_2 is $\phi_1 + \frac{1}{2}\pi$ and A_2 equals A_1, we have circular polarization along $-\hat{z}$. For the general case where A_2 and A_1 are unequal and ϕ_2 and ϕ_1 are arbitrary, the displacement $\underset{\sim}{\psi}$ describes an elliptical path. We can see this as follows: Call ψ_x and ψ_y by the names x and y. Then x is $A_1 \cos(\omega t + \phi_1)$ and y is $A_2 \cos(\omega t + \phi_2)$. Expand each of these cosines, so that x is a certain linear combination of $\cos \omega t$ and $\sin \omega t$, and y is another linear combination. Now solve these two linear equations for $\sin \omega t$ and $\cos \omega t$. You will then find that $\sin \omega t$ and $\cos \omega t$ are each a (different) linear combination of x and y. Now add the squares of $\sin \omega t$ and $\cos \omega t$. The result(which equals 1) is a quadratic expression involving x^2, y^2, and xy. Such an expression is called a conic section. If the possible values of x and y have limited magnitude (as they do here), the conic section is an ellipse. (See problem 1.) In Fig. 3 we show what happens as we vary the relative phase in Eq.(11).

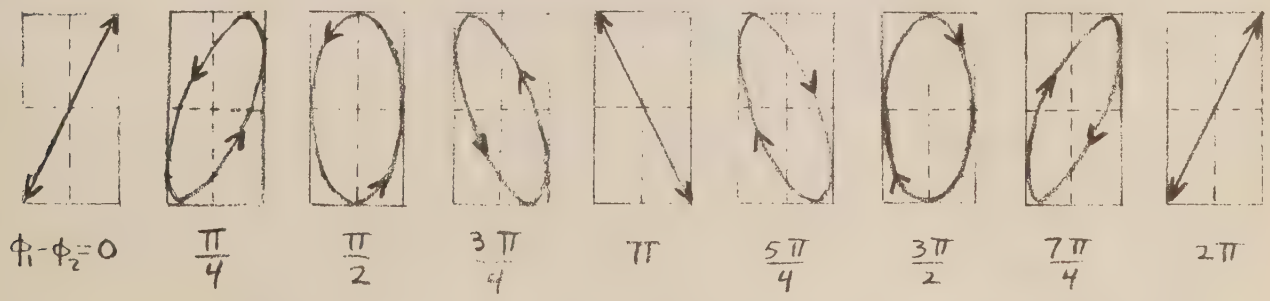


Fig. 4-3. General polarization. The amplitude for the y motion is taken to be twice that for the x motion. The y motion lags the x motion by the indicated phase constant, which is $\phi_1 - \phi_2$.

A general polarization can be expressed as a superposition of two linearly polarized components, one along $\hat{x}$ and one along $\hat{y}$, as in Eq. (11), or alternatively it may be expressed as a superposition of two circularly polarized components, one circularly polarized along $+\hat{z}$ and the other along $-\hat{z}$. Thus $\underset{\sim}{\psi}(t)$ as expressed by Eq. (11) can be written instead in the form

$$\underset{\sim}{\psi}(t) = \hat{e}_+ A_+ + \hat{e}_- A_- \quad , \tag{12}$$

where A_+ and A_- are the amplitudes for circular polarization along $+\hat{z}$ and $-\hat{z}$, and $\hat{e}_+$ and $\hat{e}_-$ are given by Eqs. (6) and (7). To relate A_1, A_2, ϕ_1, and ϕ_2 of Eq. (11) to A_+, A_-, ϕ_+, and ϕ_- of Eq. (12) involves only algebra. (See problem 2.)

Complex notation. When there are several phase constants in the problem, such as ϕ_1 and ϕ_2 or ϕ_+ and ϕ_-, it is generally easier to use complex numbers. Thus $\underset{\sim}{\psi}(t)$ as given by Eq. (11) may be taken to be the real part of a complex displacement $\underset{\sim}{\psi}_c(t)$ given by

$$\underset{\sim}{\psi}_c(t) = \hat{x}\, e^{i\omega t} A_1 e^{i\phi_1} + \hat{y}\, e^{i\omega t} A_2 e^{i\phi_2} \quad , \tag{13}$$

or in terms of circularly polarized components by

$$\underset{\sim}{\psi}_c(t) = \hat{e}_+ A_+ e^{i\phi_+} + e_- A_- e^{i\phi_-} \quad , \tag{14}$$

with

$$\hat{e}_+ = \hat{x}\, e^{i\omega t} + \hat{y}\, e^{i(\omega t - \frac{1}{2}\pi)} \quad , \tag{15}$$

$$\hat{e}_- = \hat{x}\, e^{i\omega t} + \hat{y}\, e^{i(\omega t + \frac{1}{2}\pi)} \quad . \tag{16}$$

You can easily verify that the real part of $\underset{\sim}{\psi}_c$ as given by Eqs. (14), (15), and (16) is the same as $\underset{\sim}{\psi}$ given by Eqs. (12), (6), and (7).

When you wish to include the z dependence of the waves you can

do the following (for either the real or complex wave functions): To obtain a standing wave (with a node at $z = 0$) multiply $\psi(t)$ by $\sin kz$. To obtain a traveling wave (propagating in the $+z$ direction) replace ωt by $\omega t - kz$. Thus a standing wave (with node at $z = 0$) is the real part of

$$\psi_c(z,t) = \sin kz \left[\hat{x}\, e^{i\omega t} A_1 e^{i\phi_1} + \hat{y}\, e^{i\omega t} A_2 e^{i\phi_2}\right] \tag{17}$$

$$= \sin kz \left[\hat{e}_+ A_+ e^{i\phi_+} + \hat{e}_- A_- e^{i\phi_-}\right] , \tag{18}$$

with $\hat{e}_+$ and $\hat{e}_-$ given by Eqs. (15) and (16). A traveling wave (propagating in the $+z$ direction) is given by the real part of

$$\psi_c(z,t) = e^{i(\omega t - kz)}\left[\hat{x}\, A_1 e^{i\phi_1} + \hat{y}\, A_2 e^{i\phi_2}\right] \tag{19}$$

$$= e^{i(\omega t - kz)}\left[\left(\hat{x}+\hat{y}\, e^{-\frac{1}{2}i\pi}\right) A_+ e^{i\phi_+} + \left(\hat{x}+\hat{y}\, e^{+\frac{1}{2}i\pi}\right) A_- e^{i\phi_-}\right] . \tag{20}$$

4.3 Production of polarized transverse waves.

Now that we know how to describe the polarization of harmonic waves [i.e., by giving A_1, A_2, ϕ_1, and ϕ_2 in Eq. (11) or by giving the corresponding quantities for the circular polarization representation of Eq. (12)], let us see how we can obtain a given type of polarization.

Production of desired polarization by control of the emission process. When you shake a slinky you control the polarization state of the waves by controlling the direction of the shaking. Similarly, the radio waves or microwaves emitted by an antenna have polarization which depends on the motion of the electrons in the antenna. If the antenna is a straight piece of wire normal to $\hat{z}$, the electrons oscillating along the wire shake the

electric lines of force in that direction, and the radiated electromagnetic waves are linearly polarized with the electric field parallel to the antenna. If one has one straight antenna along $\hat{x}$ and another along $\hat{y}$ and they are driven by equal currents which have the same phase, the radiation will be linearly polarized along the 45° direction between $\hat{x}$ and $\hat{y}$. If the x current is equal in amplitude to the y current but leads it by 90° in phase, the electromagnetic radiation will be circularly polarized along $+\hat{z}$. Polarization by selective absorption. If one starts with a general state of polarization as given by Eqs. (11) or (12), one way to get a given polarization is to get rid of the undesired components by arranging things so that the undesired component of the waves does work on some "moving parts," whereas the desired component does not. Consider standing waves on a slinky. Suppose $\hat{z}$ is horizontal (along the slinky), $\hat{y}$ is vertical, and $\hat{x}$ is horizontal. A vertical massless (styrofoam) push rod is attached to a massless piston that stirs the water in a bucket. The piston will be driven by the y component of vibrations. If we start out with a standing wave that includes both x and y vibrations in equal amounts, the y vibrations will soon be damped out as their energy is converted into heat in the water bucket (provided we do not keep regenerating them by shaking).

In the case of microwaves we can accomplish selective absorption by means of a skein of parallel conducting wires stretched along $\hat{y}$, as shown in Fig. 4. Suppose that the electric field in the incident electromagnetic radiation (microwave radiation) has both x and y components. We may consider the effect of the wires on these two components separately. First consider the y component, along the wires. The electric field of the incident radiation drives the electrons along the wire. The wire (if it is

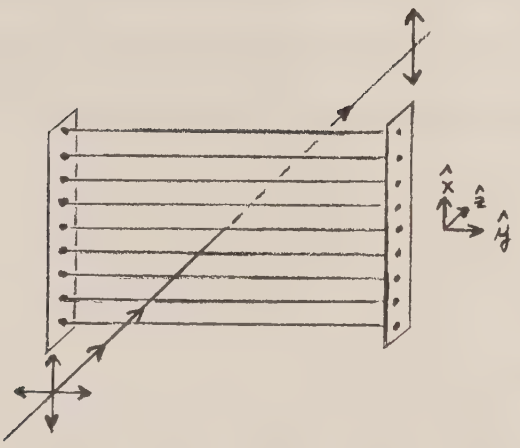


Fig. 4-4. Skein of wires absorbs those microwaves with $\underset{\sim}{E}$ along $\hat{y}$.

made of copper or silver or any good metallic conductor) acts as a resistive load. The conduction electrons reach terminal velocity in a time short compared with the period of the microwaves (which we may take as having a frequency of 1000 Mc, for example). The electric field does work on the electrons; they transfer some of their energy to the copper lattice through collisions. They also radiate. In the forward direction their radiation interferes destructively with the incident radiation and cancels it essentially to zero. In the backwards direction the radiation due to motion of the electrons driven along $\hat{y}$ gives a reflected wave. (Actually, it turns out that only a small part of the energy of the incident radiation with $\hat{E}$ along $\hat{y}$ is turned into heat in the wires. Most is reflected back in the $-\hat{z}$ direction.) Thus the skein of wires eliminates the y component.

Now consider what happens along $\hat{x}$. The electrons are not free to move along $\hat{x}$, because they cannot leave the wire. Instead of reaching a steady terminal velocity (as they do for motion along $\hat{y}$), the electrons soon build

up a surface charge along the +x and -x edges of the wire. When the field due to the surface charge is sufficient to cancel the incident field (inside the wire), the electrons stop moving. This happens in a time short compared with the microwave period. Thus the electrons are always in a sort of static equilibrium (or nearly so) with no velocity or acceleration. They do not absorb energy, nor do they radiate. Consequently the x component of the radiation is unaffected.

It may have occurred to you that surface charge will also build up on the $\pm y$ ends of the wires. However the resulting field from these end charges (which tends to cancel the incident field along $\hat{y}$ inside the wires) can be made as small as we wish in the region of interest (near the center of the skein) by making the wires sufficiently long in the y direction.

In order to completely absorb (or reflect, actually) the microwave component along $\hat{y}$, it takes a certain total number of conduction electrons, i.e., a certain number of wires to get sufficient material. The amount of material that it takes is roughly the amount which would give a sheet several "skin depths" deep if the material were spread in a uniform sheet instead of in wires.

For visible light with $\lambda \sim 5 \times 10^{-5}$ cm, it is not easy to make parallel conducting "wires" spaced at less than λ. Nevertheless it has been done!*

In 1938 Alfred Land invented polaroid, which behaves somewhat like a wire grid. In the manufacturing process, a sheet of plastic consisting of long hydrocarbon chains is greatly stretched in one direction. This lines up the molecules. Next, the sheet is dipped in a solution containing iodine.

---

*G. R. Bird and M. Parrish, Jr., J. Opt. Soc. Am. 50, 886 (1960) evaporated gold at a glancing angle onto a plastic diffraction grating having 50,000 parallel scratches per inch. The gold deposited on the downstream sides of the scratches to form parallel conducting "wires."

The iodine attaches to the long hydrocarbon chains and provides conduction electrons that can move along the chains but not perpendicular to them. One thus has effective "wires." Sometimes one calls polaroid an example of a dichroic filter.

A sheet of polaroid thus has an axis (lying in the sheet) called the axis of easy transmission. If $\underset{\sim}{E}$ is along this axis the light is transmitted with very little absorption. If $\underset{\sim}{E}$ is perpendicular to the easy-transmission axis the light is almost completely absorbed. The easy-transmission axis is perpendicular to the direction of stretching of the plastic, i.e., it is perpendicular to the "wires."

When you look at a white sheet of paper through a piece of polaroid, the paper looks grey. That is because half of the light coming from the paper is absorbed by the polaroid. There is only half as much light, so it naturally looks darker. On the other hand, a piece of clear cellophane (or other clear plastic) transmits almost all of the light incident on it.

If light is incident along $\hat{z}$ with electric field amplitude $\underset{\sim}{E}$ and if $\hat{e}$ is the direction of easy transmission for the polaroid, then only the amplitude component $\underset{\sim}{E} \cdot \hat{e}$ is transmitted. The transmitted energy flux I_{out} is less than the incident flux I_{in} by a factor $(\underset{\sim}{E} \cdot \hat{e})^2/(\underset{\sim}{E}^2)$:

$$I_{out} = I_{in} \cos^2 \theta \equiv I_{in} (\hat{\underset{\sim}{E}} \cdot \hat{e})^2 \qquad (21)$$

Equation (21) is often called Malus's law. See Fig. 5.

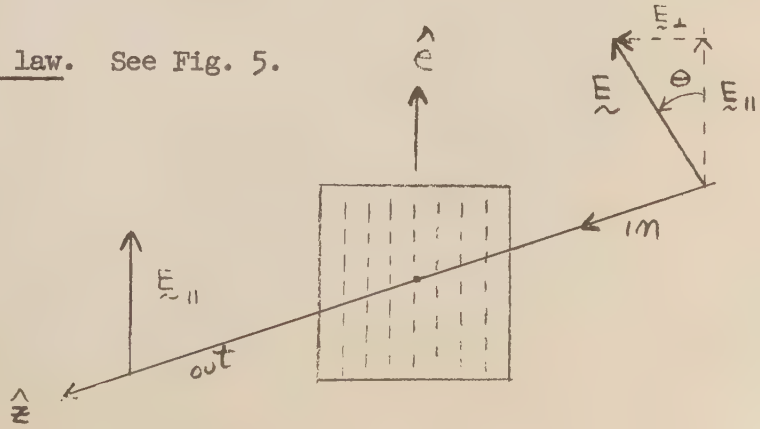


Fig. 4-5. Polaroid. The axis for easy transmission of $\underset{\sim}{E}$ is along $\hat{e}$. The component of $\underset{\sim}{E}$ that is parallel to $\hat{e}$, $E_{\parallel}$, is transmitted. The other component, $\underset{\sim}{E}_{\perp}$, is completely absorbed.

Two successive polaroids Nos. 1 and 2 with easy axes $\hat{e}_1$ and $\hat{e}_2$ at 90 degrees to one another are called "crossed" polaroids. The first polaroid transmits $\underset{\sim}{E}$ along $\hat{e}_1$ and the second polaroid completely absorbs this field. No light is transmitted beyond the second polaroid. However, if a third polaroid is inserted between the crossed polaroids the transmitted field is not zero, provided $\hat{e}_3$ is along neither $\hat{e}_1$ nor $\hat{e}_2$. See problem 3. Also, you can prove this with the pieces of polaroid from your optics kit. (We only gave you two pieces. Cut off a quarter-inch-wide piece from one edge of one piece. We suggest cutting it so that the remaining oblong piece has its longest dimension along its axis of easy transmission, so as to provide a convenient reference "pointer." You can find that axis by looking at a source having known polarization. For example, look at light reflected from a window or from a wooden or plastic floor. The reflected light is polarized parallel to the flat surface, say the floor. Make the long edge be the one parallel to the floor when the transmitted light is brightest.)

Polarization by single scattering. On a sunny day look at the blue sky through a piece of polaroid. Hold the polaroid close to one eye so that you can see a large area of sky. Rotate the polaroid and look for a minimum that looks like a dark swath across the sky. The light coming from that part of the sky is strongly polarized. Measure (roughly) the angle between the line connecting your head and the sun and the line connecting your head and the region of greatest polarization of the blue sky. (You will find it is about 90°). Measure the direction of polarization.

The explanation for the polarization of the blue sky is as follows: Let $\hat{z}_1$ be the direction of propagation of light from the sun to a given

air molecule. The electric field in the sunlight is unpolarized. (You may verify this by cutting a small hole in a piece of cardboard, holding the cardboard so that the sunlight makes a bright spot on the floor where it comes through the hole, covering the hole with a piece of polaroid, rotating the polaroid and looking for a variation in the intensity of the spot on the floor. Don't look at the sun!) The electrons in an air molecule act somewhat like a driven oscillator driven by the incident light. They therefore oscillate in a superposition of motions along $\hat{x}_1$ and $\hat{y}_1$ (the directions transverse to $\hat{z}_1$). The oscillating electrons radiate in all directions, but they do not radiate equally well in all directions. It turns out that the amplitude and polarization direction of the electric field radiated by a single oscillating point charge are proportional to the "projected" amplitude of motion of the oscillating charge, as seen by an observer who looks toward the radiating oscillating charge. By the projected amplitude of motion we mean the amplitude of that vector component of the electron's motion which is perpendicular to the propagation direction $\hat{z}_2$ from the oscillating charge to the observer. If $\hat{z}_2$ is perpendicular to $\hat{z}_1$, then (taking $\hat{y}_1$ along $\hat{z}_2$ and $\hat{x}_1$ transverse to $\hat{z}_2$) The observer only sees the $\hat{x}_1$ component of the electron's motion. He therefore observes radiation which is 100% linearly polarized along $\hat{x}_1$. The intensity is only half what it would be if he were looking back along the $\hat{z}_1$ axis and could see both the $\hat{x}_1$ and $\hat{y}_1$ motions of the electron. [In our example it is difficult to look back directly along $\hat{z}_1$ because then we are blinded by the sun. But you can look at various angles and see that if you look near the sun the sky appears unpolarized. That is also the case if you look at air that requires large scattering angles (as near 180° as you can manage) to get to your eye.]

This polarization process is illustrated in Fig. 6.

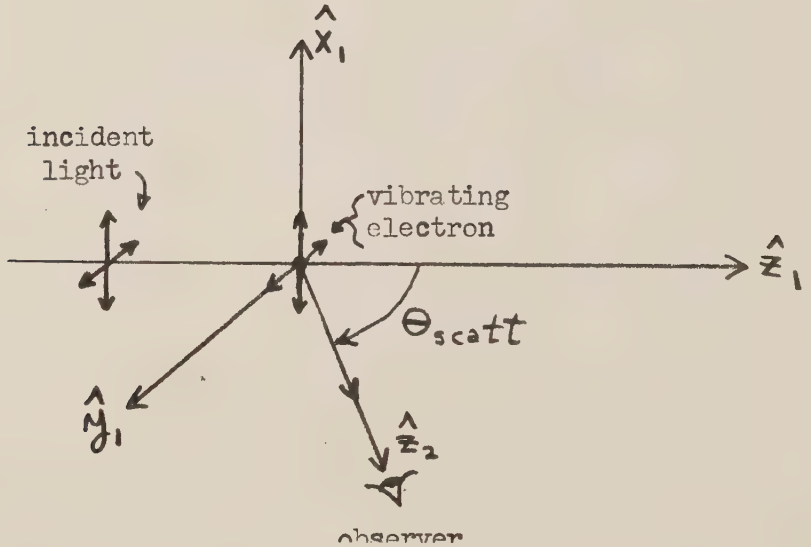


Fig. 4-6. Polarization by single scattering. The $\hat{y}_1$ axis is chosen to lie in the plane of $\hat{z}_1$ and $\hat{z}_2$. The observer "sees" the full projected motion of the electron along $\hat{x}_1$, but he sees a projected amplitude along $\hat{y}_1$ that is only a fraction $\cos \theta_{scatt}$ of the actual $\hat{y}_1$ motion. At $\theta_{scatt} = 90^\circ$, the scattered radiation is 100% polarized along $\hat{x}_1$.

Bees don't need polaroids to detect the polarization of the blue sky. They use it to navigate.*

*Bees, Their Vision, Chemical Sense, and Language, Karl von Frisch, Cornell University Press (1950).

Depolarization by multiple scattering. The bluish reflected light by which we see a searchlight beam "from the side" as it penetrates ordinary (i.e., smoggy) air is polarized by the same mechanism as is the blue sky. If the air is foggy the searchlight beam appears white instead of bluish; then the light is not polarized. Neither is sunlight polarized when it is reflected from white clouds, or from sugar, or from a white piece of paper. Although a single scatter at just the right angle can give strongly polarized scattered light, that does not mean that many scatters are even better! If you reflect light at the correct angle from a piece of glass, you get 100% linearly polarized light. (This is discussed in the next paragraph.) But if you grind up the glass into a powder, the light that comes to your eye from a given part of the surface will have suffered many reflections and will have penetrated to considerable depth into the ground glass. As a result you are looking at electrons vibrating in all directions (transverse to your line of sight), since they have been excited by radiation incident on them from many directions besides the original one from the light source. They are even being driven by light that has penetrated a little ways, has been reflected several times, and is on its way back out. You can obtain a nice demonstration of the depolarizing effect of multiple scattering by inserting a piece of ordinary translucent waxed paper between two crossed polaroids. The waxed paper almost completely depolarizes the polarized light produced by the first polaroid. The fact that the waxed paper is multiply scattering the light can be shown by the simple experiment of laying the waxed paper on a printed page. If the paper is very close to the print you can easily see the black letters. If the waxed paper is moved

an inch above the page, the letters become so blurred that they cannot be distinguished. Think of the "black light" from a letter on the page to your eye as a little flashlight beam. The beam gets diffused by the waxed paper. (Another good experiment is to take a flashlight beam and shine it through the waxed paper. Look at the size of the transmitted spot as you move the flashlight farther and farther behind the waxed paper.) A clear piece of glass or plastic does not multiply scatter incident light (you can read through it whether or not it is close to the print) and does not depolarize.

Polarization by specular reflection. Brewster's angle. Look at the reflection of something in a piece of ordinary window glass or from a smooth surface of water. Test the polarization of the reflected light with a piece of polaroid. You will find that at an angle of incidence of about 56° (measured from the incident ray to the normal to the surface) for glass of index $n = 1.5$, or at an angle of incidence of about 53° for water (index about 1.33), the reflected light is 100% linearly polarized parallel to the surface. This special angle is called Brewster's angle. By rotating the polaroid to the proper position you can completely extinguish the reflected light, if it is incident at Brewster's angle. (If you properly orient the polaroid and hold it close to one eye so as to see a broad range of angles, you will see a band of "extinction" centered at Brewster's angle.)

For any angle of incidence, the angles of the incident and refracted rays, namely θ_1 and θ_2, are related by Snell's law (see Chap. 3, p.):

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \ . \qquad (22)$$

The incident and reflected rays make equal angles with the normal. (This is called the law of specular reflection.) Thus the particular angle of incidence θ_1 for which $\theta_1 + \theta_2$ is 90°, the reflected ray makes an angle of 90 degrees with the refracted (i.e., transmitted) ray, as shown in Fig. 7.

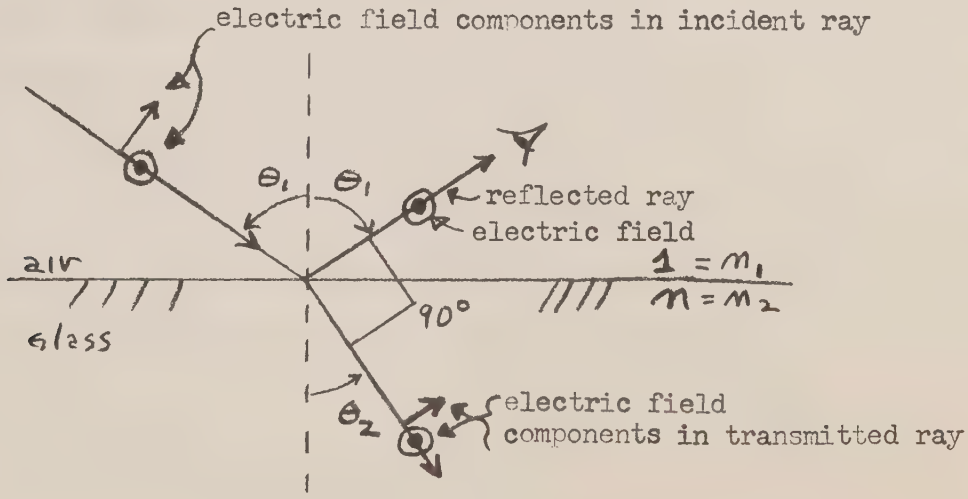


Fig. 4-7. Brewster's angle. The angles are drawn correctly for glass (n = 1.5). The reflected light is 100% polarized perpendicular to the plane of incidence, i.e., the plane of the incident ray and the normal. (The circled dot indicates electric field polarization into or out of the paper.)

The direction of oscillation of the electrons in the glass is transverse to the direction of the transmitted ray (since that is the direction of their driving force). For any angle of incidence, the component of electron motion perpendicular to the plane of incidence (perpendicular to the plane of the paper in Fig. 7) is completely "visible" to an observer looking at the reflected light (the light radiated by the driven electrons in the glass), since this component

of the motion is perpendicular to the direction of propagation from electron to observer (i.e., the direction of the reflected ray). However, the component of electron motion that lies in the plane of incidence is not perpendicular to the direction of the reflected ray. Only the component of motion projected perpendicular to the reflected ray contributes to the reflected radiation. At Brewster's angle of incidence, the motion of the electrons in the plane of incidence is exactly along the line from the electron to the observer and contributes no reflected light. Thus the reflected light is completely polarized perpendicular to the plane of incidence. From Fig. 7 we see that this condition corresponds to $\theta_1 + \theta_2$ equal to 90°. Therefore Eq. (22) gives [using $n_1 = 1$, $n_2 = n$, and $\sin\theta_2$ equal to $\sin(90^\circ - \theta_1)$, which is $\cos\theta_1$]

$$\tan\theta_1 = n \quad , \quad \theta_1 = \text{Brewster's angle.} \qquad (23)$$

The phase relations between the incident, transmitted, and reflected light are interesting. They are shown in Fig. 8, where a dot means $\underset{\sim}{E}$ is out of the paper, a cross means $\underset{\sim}{E}$ is into the paper, and an arrow in the plane of the paper indicates $\underset{\sim}{E}$ along the arrow.

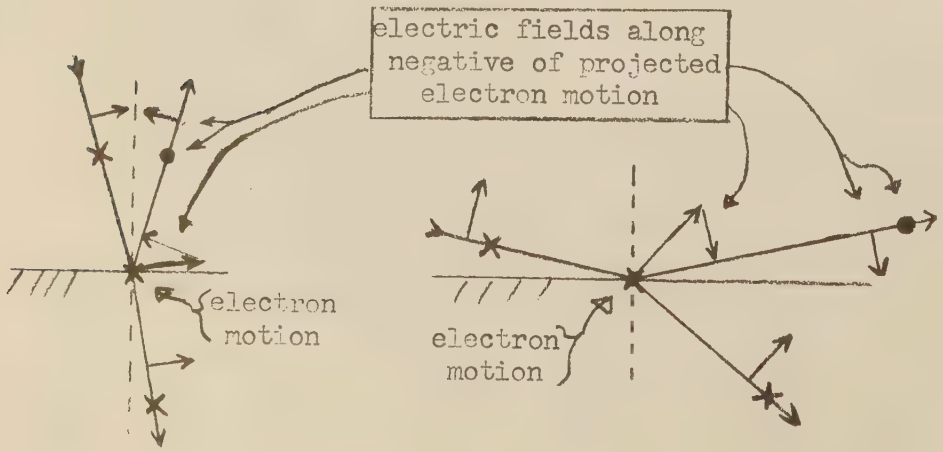


Fig. 4-8. Phase relations in light reflected from glass.

(a) θ_1 less than Brewster's angle

(b) θ_1 greater than Brewster's angle

The phase relations can be understood as follows. The transmitted wave (which is partly due to the original light source and partly due to the radiation of the driven electrons) always has the same phase as the incident wave. (In Chap. we found the transmission coefficient to be $1 + R_{12}$, with R_{12} lying between -1 and +1. Thus the transmission coefficient is always positive.) Therefore we can draw in the arrow and the cross on the transmitted ray in both parts of Fig. 8. The electrons are driven by the fields in the transmitted ray. Therefore the electron motion corresponding to a given transmitted polarization is along that polarization. Let us therefore take the cross and arrow giving the transmitted polarization [in Figs. 8(a) and 8(b)], redraw them at the point of intersection of the rays at the surface, and say that they now represent the motion of the driven electrons. This motion produces a reflected wave with electric field along the negative of the projected motion of the electron. (That is what we learned in Chap. 3 for the case of normal incidence, where we found $R_E = (n_1-n_2)/(n_1+n_2) = -0.2$ for light incident from air to glass.) You can easily verify the phase relations exhibited in Fig. 8 for yourself. (See Home Experiment 20.)

One interesting application of Brewster's angle is in the design of a glass window having 100 percent transmission, called a Brewster window. Suppose you have an apparatus in which it is necessary or convenient to send a light beam through a glass window. At normal incidence one transmits only about 92 percent in intensity through a glass window. (About 4 percent is lost at each surface.) This may be tolerable in some situations, but not in a gas LASER with mirrors outside the windows, where one wants perhaps 100 traversals of the window. An ingenious solution is

to work only with polarized light and then tilt the window so that the light is polarized parallel to the surface and is incident at Brewster's angle. Then one has zero reflectance at either the first or second surface, and the light is completely transmitted. Thus instead of losing almost all the light of both polarizations, as you would with 100 traversals at normal incidence (note that 0.92 to the hundredth power is about 1/4000), you lose all of one polarization but none of the other. See Fig. 9.

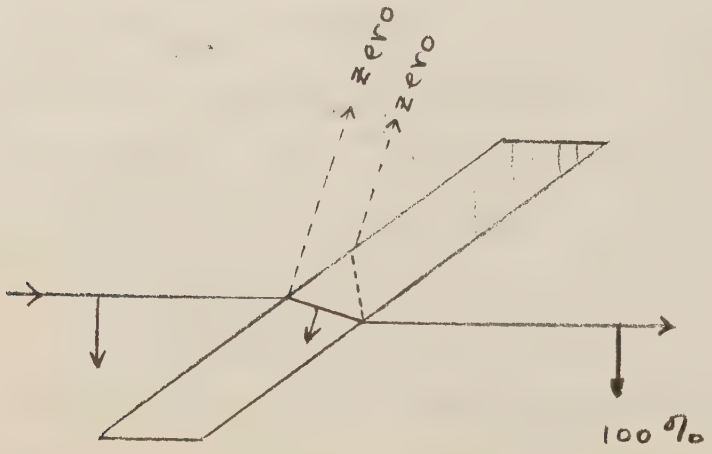


Fig. 4-9. Brewster window. The figure is drawn for $n = 1.5$.

Even more spectacular than the polarization of the blue sky is the polarization of the rainbow. It is an interesting exercise to try to predict whether the polarization is radial or tangential (with respect to the bow). If you cannot wait for rain to verify your prediction, get a garden hose and wait for sun (or use a source of light at night). For an explanation of how a rainbow works, we recommend PSSC Physics, 1st edition, p. 222. (We couldn't find it in the second edition.)

4.4. Double refraction.

In Sec. 4.3 we learned how to change the state of polarization of a beam of electromagnetic waves by means of selective absorption or by selective reflection (selective in the sense that one polarization component is absorbed or reflected more than the other). In this section we shall learn how to change the state of polarization by changing the relative phase of the two components.

Take two pieces of polaroid and cross them so that no light gets through. Now slip an ordinary piece of cellophane (from a candy wrapper or bread wrapper or something -- almost any kind of clear plastic will do) between the crossed polaroids. Now light gets through! Since the cellophane is perfectly transparent, with none of the "dark" appearance of polaroid, we know that the cellophane cannot be absorbing light. The only way it can change the polarization of the light is to change the relative phases of different polarization components. (Then there will be no loss of intensity, as you can easily show.)

Now rotate the piece of cellophane between the polaroids, keeping the polaroids crossed. You should find two angles at 90° from one another (in 180° of rotation) where the cellophane has its greatest effect and two angles at 90° from one another where the cellophane has no effect. Thus cellophane has two special directions, oriented at 90° to one another, and lying in the plane of the cellophane, which are related to the property of introducing relative phase shifts in different polarization components of light.

Just to show that not all transparent plastic has this peculiar property, find a piece of household "Saranwrap" (there are many other brands) or a piece of the stretchy polyethylene used by dry cleaners to keep dust off clothes. Try this between crossed polaroids. You will find that it has little effect. Not much light gets through. (There will be some effect, but it will be small

compared with that of the cellophane.) Since (or if) you have now found a piece of plastic that has no "optic axes," i.e., no special directions in the plane of the sheet, try to create a special direction. Take a piece of this stretchy plastic (like Saranwrap) and stretch it. Lay the plastic between the crossed polaroids with the stretch direction at 45° to the axes of the crossed polaroids. You should now get a huge effect.

Here is the explanation for the behavior of stretched Saranwrap. Before the stretching, the long organic molecules of the plastic were coiled in a spaghetti-like mess going in all directions. However, the molecules tend to get straightened out and aligned by the stretching. The electrons in a single long chain-like organic molecule have different "effective spring constants" for vibrations along the hydrocarbon chain as compared with vibrations in the two directions perpendicular to it. Therefore the polarizability of the molecule is different for displacements along and perpendicular to the hydrocarbon chain. After the stretching operation, the direction of the long dimension of the molecules tends to lie along the stretching direction. One of the directions perpendicular to the stretching direction can be taken to lie in the plane of the sheet of plastic. (The other is perpendicular to the sheet and does not concern us.) The electric susceptibility (induced polarization per unit volume and per unit incident electric field) for electric fields along the stretch direction will thus be different from that for fields perpendicular to the stretch direction. Thus the dielectric constant is different for these two directions, and thus the index of refraction is different for these two directions.

These two directions, the stretch direction and the direction perpendicular to it (and lying in the plane of the sheet), are called the

optic axes. The optic axis yielding the larger of the two indices of refraction (for $\underset{\sim}{E}$ directed along it) is called the slow axis. The other optic axis is called the fast axis. (Larger index of refraction means slower phase velocity.) We will call the two corresponding indices of refraction n_f and n_s (f stands for fast, s for slow), with $n_s > n_f$. A sheet of cellophane or plastic or other material having these properties is called a retardation plate.

Now let us examine the effect of a retardation plate on an incident traveling electromagnetic plane wave. First let us resolve the incident electric field into orthogonal components along the slow axis $\hat{e}_s$ and the fast axis $\hat{e}_f$. Suppose that there is vacuum for $z < 0$, and the retardation plate begins at $z = 0$ and extends to $z = \Delta z$, after which we again have vacuum. Suppose that the oscillations of the electric field of the incident wave at $z = 0$ are given by the real part of the complex quantity

$$E_c(0,t) = e^{i\omega t}\left[\hat{e}_s A_s e^{i\phi_s} + \hat{e}_f A_f e^{i\phi_f}\right] . \qquad (24)$$

The amplitudes A_s and A_f and phase constants ϕ_s and ϕ_f are those we get when we resolve the incident electric field into linearly polarized components along $\hat{e}_s$ and $\hat{e}_f$. (Since these amplitudes and phase constants are arbitrary, Eq. (24) represents a general polarization.) Now consider the transmitted wave inside the retardation plate between $z = 0$ and Δz. We neglect any loss due to reflection at the first surface and then merely replace ωt by $\omega t - kz$ in Eq.(24). But we must remember that k is not the same for $\underset{\sim}{E}$ along $\hat{e}_s$ as it is for $\underset{\sim}{E}$ along $\hat{e}_f$. Thus, recalling that k is proportional to the index of refraction (big index means slow

velocity, hence small ω/k , hence large k) and is given by $n\,\omega/c$, we have inside the retardation plate

$$E_c(z,t) = e^{i\omega t}\left[\hat{e}_s A_s e^{i\phi_s} e^{-in_s\,\omega z/c} + \hat{e}_f A_f e^{i\phi_f} e^{-in_f\omega z/c}\right] . \qquad (25)$$

By the time the wave has reached the exit of the plate at $z = \Delta z$, each component has suffered a phase retardation relative to the phase it would have had if the plate had been vacuum (with $n = 1$). For the s component, this is given by $(n_s - 1)\,\omega\Delta z/c$:

$$\text{phase retardation of } E_s \text{ relative to vacuum} = (n_s-1)\,\frac{\omega\Delta z}{c} . \qquad (26)$$

Similarly, we have

$$\text{phase retardation of } E_f \text{ relative to vacuum} = (n_f-1)\,\frac{\omega\Delta z}{c} . \qquad (27)$$

Subtracting Eq. (27) from Eq. (26) we find the retardation in phase of E_s relative to E_f :

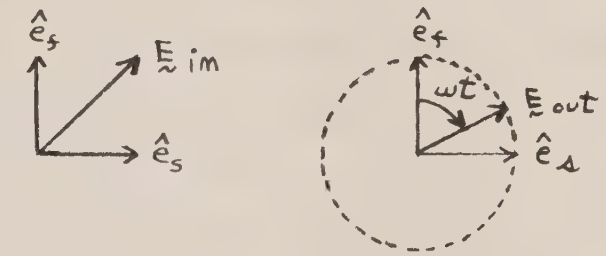
$$\text{phase retardation of } E_s \text{ relative to } E_f = (n_s-n_f)\,\frac{\omega\Delta z}{c} \qquad (28)$$

$$= (n_s-n_f)2\pi\,\frac{\Delta z}{\lambda_{vac.}}$$

where λ_{vac} is the wavelength in vacuum.

Consider the following example, which should help you keep the signs straight. Suppose incident linearly polarized light has $\underset{\sim}{E}$ along the 45° line between $\hat{e}_s$ and $\hat{e}_f$. Then A_s and A_f are equal, and ϕ_s and ϕ_f are equal. Suppose the thickness of the plate is such that the slow component suffers a retardation of $\frac{1}{4}$ cycle relative to the fast component, i.e., it suffers a phase retardation of $\pi/2$ relative to the fast component. Such a retardation plate is called a quarter-wave plate. The wave emerging at the

rear of the plate has equal amplitudes for the slow and fast components, and the fast component leads the slow by 90° in phase. That means that we have circularly polarized light with the sense of rotation from $\hat{e}_f$ towards $\hat{e}_s$. This is shown in Fig. 10.



(a) Input. (b) Output.

Fig. 4-10. Quarter-wave plate. Incident linearly polarized light with $\underset{\sim}{E}$ at 45° to optic axis.

You should convince yourself of the truth of the following numbered statements and "rules." (You should not memorize them. You should understand them well enough so that you can forget the answers and figure them out whenever you want to.)

1. A half-wave plate (one twice as thick as a quarter wave plate) converts linearly polarized light into linearly polarized light with the direction of polarization of the output obtained from that of the input by reflection in one of the optic axes. (We are almost never concerned with which axis, i.e., with the absolute phase. We don't care about a minus sign in the direction of the amplitude.) That is, a half-wave plate reverses the relative sign of the linear components of the incident amplitude.

2. A half-wave plate converts right-handed circularly polarized light into left-handed circularly polarized light and vice versa.

3. A quarter-wave plate converts linearly polarized light with polarization lying somewhere between $\hat{e}_s$ and $\hat{e}_f$ into elliptically polarized light with sense of rotation from $\hat{e}_f$ to $\hat{e}_s$. If the incident polarization is at 45° to $\hat{e}_s$ and to $\hat{e}_f$, the output polarization is circular. (Notice that this means that if we rotate the linear polarization of $\underset{\sim}{E}_{in}$ in Fig.10 by 90°, the output will rotate at frequency ω in the opposite sense to that shown in Fig. 10. To use the "rule" to see this, simply reverse the sign convention on one or the other of $\hat{e}_f$ or $\hat{e}_s$ in Fig. 10, so that $\underset{\sim}{E}_{in}$ will again lie between the $\hat{e}_s$ and $\hat{e}_f$ unit vectors. Then the rule says that the rotation is from $\hat{e}_f$ to $\hat{e}_s$.)

4. A quarter-wave plate converts circularly polarized light into linearly polarized light. To obtain a simple rule, label the signs of the slow and fast axes so that the sense of rotation is from fast axis to slow axis. Then the quarter-wave plate converts the circularly polarized light into linear polarization at 90° to the direction lying midway between $\hat{e}_s$ and $\hat{e}_f$. (The f vibration already led in phase by $\frac{1}{4}$ cycle. After the quarter-wave plate it leads by $\frac{1}{2}$ cycle.)

5. A retardation plate has no influence on the state of polarization of linearly polarized incident light with $\underset{\sim}{E}$ along either $\hat{e}_s$ or $\hat{e}_f$.

6. A retardation plate cannot convert "unpolarized" light (the kind you get directly from a light bulb or the sun) into polarized light. We shall study unpolarized light in Sec. 4.5. For now, we just say vaguely that for unpolarized light there is a "random" phase relation between the x and y components, when you average over the observation time interval. The relative phase shift introduced by the retardation plate still leaves the x and y phases as random as before -- if ϕ_x and ϕ_y are "randomly" related, so are ϕ_x and $\phi_y + \Delta\phi$.

7. A circular polarizer is obtained if we make a "sandwich" by gluing together (face to face) a piece of polaroid and a piece of quarter-wave plate with its optic axes at 45° to the easy-transmission direction of the polaroid. The sense of rotation is from the fast axis to the easy axis of the polaroid. The unpolarized light must be incident on the polaroid face of the sandwich.

8. A circular polarizer that produces right-handed circularly polarized light will transmit with 100% efficiency (neglecting small losses by reflection) right-handed circularly polarized light traveling in the reverse direction (i.e., incident on the quarter-wave plate face of the sandwich.) It will completely absorb left-handed circularly polarized light incident on the quarter-wave plate face. (This fact is easily remembered by analogy with a die and a screw. A die that converts a cylindrical "unpolarized" rod into a right-handed screw will also"transmit" a right-handed screw in the reverse direction, but it will completely annihilate a left-handed screw traveling in the reverse direction.) This fact has interesting consequences. See Home Experiment 22.

We have been considering retardation plates made by stretching a sheet of plastic in one direction. That is what (we hope) you did with a piece of Saranwrap. That is the way Polaroid Corporation makes the quarter-wave and half-wave plates that are included in your optics kit. That is the way cellophane acquires its properties (it comes out between rollers that squeeze it and line up the molecules). An ordinary piece of window glass is isotropic and exhibits no double refraction (i.e., has no optic axis). But if you look at a piece of stressed plate glass between crossed polaroids, you will see transmitted light in some places. Safety glasses are highly stressed and exhibit interesting patterns of double refraction. Plastic triangles and dishes show beautiful colored stress patterns when put between crossed polaroids.

The color effects are partly due to the variation of the index of refraction with color, but mostly they are due to the variation of the phase shift with wavelength, assuming n is constant.

Most crystalline materials exhibit double refraction. If they have only one direction of anisotropy (like the stretched plastic has) they are called uniaxial. The direction along the axis of anisotropy is called the "extraordinary" direction. The other two directions perpendicular to the anisotropy axis are called "ordinary" directions. The corresponding indices are called n_e and n_o (e for extraordinary and o for ordinary). The anisotropy axis can be either a fast axis or a slow axis, depending on the crystal structure. Here are some examples, with indices of refraction for light of wavelength 5890A° (yellow light emitted by excited sodium atoms):

Material	n_e	n_o	e axis
Quartz	1.553	1.544	slow
Calcite	1.486	1.658	fast
Ice	1.307	1.306	slow

Optical activity. Here is an interesting home experiment. Fill a glass-bottomed jar or glass (not a plastic one) with about two inches of Karo corn syrup (obtained at any grocery store). Put a light source under the jar, a piece of polaroid under the jar, and one above it. Now look through the syrup. You will see beautiful color effects. Now investigate quantitatively: Use red or blue cellophane (obtained at any variety store or art store) for a filter in order to work with a reasonably small band of wavelengths. (You may look at a light bulb using your Edmund diffraction grating from the optics kit with the cellophane either in or out to see what sort of band of colors you

are using.) Vary the depth of the syrup. You should find that the linearly polarized light remains linearly polarized, but the direction of polarization is rotated by about 30° clockwise (as you face the light) per inch of syrup.

Here is the explanation. The linearly polarized light produced by the first polaroid is a superposition of equal amounts of right-and left-handed circularly polarized light:

$$E_o \hat{x}\, e^{i\omega t} = \frac{E_o}{2}\left[\hat{x}\, e^{i\omega t} + \hat{y}\, e^{i(\omega t - \frac{1}{2}\pi)}\right] + \frac{E_o}{2}\left[\hat{x} e^{i\omega t} + \hat{y}\, e^{i(\omega t + \frac{1}{2}\pi)}\right] . \quad (30)$$

See Fig. 11.

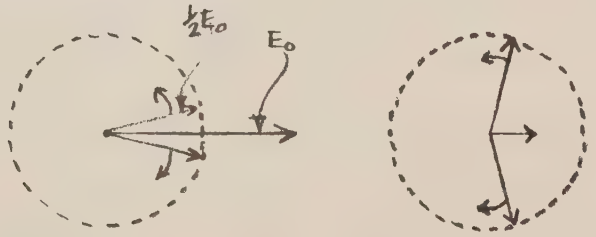


Fig. 4-11. Linearly polarized oscillation of amplitude E_o is a superposition of left-and right-handed circularly polarized oscillations each having amplitude $\frac{1}{2} E_o$. The direction of the linearly polarized superposition depends on the relative phases of the circular components.

The sugar molecules have a helical structure. All sugar molecules made from corn have the same handedness. A helix has the same handedness whether viewed from one end or the other. Therefore a solution of randomly oriented sugar molecules has a net handedness like that of the individual molecules. Because of the helical structure of the molecules, the sugar solution has

different indices of refraction for traveling waves with right-handed and with left-handed circular polarization. As the linearly polarized wave progresses through the sugar solution, one of the circularly polarized components gets ahead of the other in phase. A little thought and a sketch should convince you that the direction of rotation of the linear polarization is the same as the direction of rotation of the fast circular component (the fast one being the one with the smaller index of refraction). Here is food for thought: What happens if we send light through a sugar solution, reflect it from a mirror and send it back through the solution in the opposite direction? Does the rotation get doubled? Or is it returned to zero?

One of Louis Pasteur's first great discoveries was that racemic acid, an optically inactive form of tartaric acid, was an equal mixture of left-handed and right-handed tartaric acid. He succeeded in recognizing under a microscope the right-handed and left-handed crystals in the racemic mixture, and he separated the crystals into two piles using a fine tweezer. When dissolved in water, one pile rotated the plane of polarized light in the same direction as does natural tartaric acid produced by grapes. The other pile rotated the polarization by an equal amount in the opposite direction. That kind of tartaric acid had never been seen in the world before!*

The observed single-handedness of helical organic molecules made by present-day living organisms is undoubtedly a fundamental clue for the unraveling of the history of the evolution of life on our planet. All present DNA molecules (the stuff of life) are right-handed helixes! Why? Because of original chance? Did the oceans once contain equal amounts of right-handed and left-handed primitive DNA? Did the right-handed DNA learn

---

* For an account of this and other great experiments of Pasteur see Pasteur and Modern Science by Rene Dubos, Anchor, (1960).

how to eat up the left-handed kind? Nobody knows, yet.*

Metallic reflection. After one has observed the strong polarization (which is 100% at Brewster's angle) obtained by specular reflection from a dielectric material like glass or water, it comes as something of a shock when one looks for polarization upon reflection from ordinary aluminized or silvered mirrors, or any other silvery-looking substances, for example automobile chrome trim or a table knife. There is essentially zero polarization. That is because a silvery-looking metal reflects both polarizations almost completely. That is why it is silvery looking; it would be darker if it were polarizing by reflecting less light of one polarization than another. (To see this, place a silvery mirror beside a piece of glass and look at both near Brewster's angle for the glass. Put something dark under the glass.)

The fact that a shiny piece of metal does not produce polarized light from unpolarized light should not lead us too hastily to believe that it has no effect on polarized light. After all, a piece of cellophane does not produce polarized light from unpolarized incident light, but it can change the polarization state of incident polarized light. That is exactly what a piece of shiny metal does. You can verify this in an easy home experiment by which you convert linearly polarized light into circularly polarized light by metallic reflection. See Home Experiment 26.

---

* For a beautiful account of the role of handedness in living organisms and in the weak decay interactions of elementary particles, see The Ambidextrous Universe, Martin Gardner, Basic Books (1964).

4.5 Bandwidth, coherence time, and polarization.

In this section we shall discuss the polarization of light emitted by atoms. We shall use a classical picture of an electron bound to a heavy nucleus. In this picture, the electron oscillates and emits classical electromagnetic waves, just as if the atom were a little radio antenna. This classical picture omits the "graininess" of the emission of light, i.e., it omits the fact that light is emitted and absorbed in "lumps" called photons . Aside from that, the classical picture gives many of the same results as does sophisticated quantum theory. The main difference is that in the classical theory we think of electromagnetic waves as carrying a continuous energy flux,whereas in quantum theory we learn that the energy flux is not continuous. However, Maxwell's equations (the equations of classical electromagnetic theory) do give the right predictions for the average energy flux. Classically we think of the electric and magnetic fields of the electromagnetic radiation as being quite "tangible", with their squares giving the "actual" energy density in the wave. Quantum theory reinterprets the classical energy density as being the average number of photons times the energy of one photon. (When the average number of photons in a given volume is less than unity, it is called the probability of finding one photon.) You will study quantum theory in Volume V. We only give you these remarks so as to assure you that the results we shall obtain with a classical picture still hold in quantum theory, after suitable reinterpretation of energy flux as probability flux times photon energy.

Let us consider a single classical atom situated at $x = y = z = 0$. The electron may be oscillating in a superposition of motions along $\hat{x}$, $\hat{y}$, and $\hat{z}$. The observer of the emitted light is situated somewhere far away on the positive

z axis. Only the x and y components of the electron's motion contribute to the electromagnetic waves observed (the light).

Suppose that at $t = 0$ the electron is excited into vibration, perhaps by a collision. After $t = 0$ the electron vibrates freely at a natural frequency ω_o. The polarization state of the emitted radiation depends on the amplitudes of the x and y components of the motion and on the relative phase of the x and y motions. The electron does not vibrate forever. It loses energy by radiation, with a mean decay time τ (the time for the energy to decrease by a factor e, also called mean life). After several mean lives the electron has lost most of its energy and its radiation is negligible. During the time (of order τ) that it radiates, its x and y motions retain constant relative phase. (They both oscillate with the same frequency, and we assume the atom is undisturbed during this interval.) Therefore the polarization of the emitted radiation is constant during this time interval.

At a later time the atom may suffer a second collision, again exciting the electron into a motion that is a superposition of oscillations along $\hat{x}$, $\hat{y}$, and $\hat{z}$, all with the same natural frequency ω_o, and with amplitudes and phase constants that depend on the circumstances of the collision. If the atom is in a gas and is being bombarded uniformly from all sides, we may assume that there is practically no correlation between the x and y amplitudes and phases in successive excitations. Therefore the polarization state of the radiation emitted during the second time interval (of order τ) following the second excitation bears no relation to that emitted following the first excitation.

Now suppose that instead of one atom we have many atoms excited at any one instant. They are all located in a small region near $x = y = z = 0$, and

the observer out on the z axis sees an electromagnetic wave that is a superposition of the waves emitted by the individual atoms. Let us call an "instant" a time interval which is short compared with the mean decay time τ, but which nevertheless contains many oscillations at frequency ω_o. Suppose the observer describes the radiation in terms of the amplitudes for E_x and for E_y and the relative phase between E_x and E_y. At any "instant" E_x is a superposition of contributions from all the atoms radiating at that instant. The same is true of E_y. All atoms oscillate with the same frequency ω_o, but with different amplitudes and phase constants. Therefore the superposition E_x has frequency ω_o and has amplitude and phase constant that depend on the amplitudes and phase constants of all the contributing atoms. (The same statement holds for E_y.) During any time interval that is short compared with τ, all the atoms that are vibrating lose only a small fraction of their energy, and they maintain the same phase constants. Therefore the amplitude and phase constant of the superposition that gives E_x (or E_y) do not change substantially during a time interval short compared with τ. The polarization state of the total electromagnetic wave remains constant during a time interval short compared with τ. In particular, the relative phase of E_x and E_y remains constant. Now suppose we wait for several mean lives τ, and then reexamine the state of polarization of the total wave. After a long interval of many mean lives, the atoms that were radiating (at the beginning of the interval) have decayed to zero, and have been replaced by new atoms. (It does not matter what fraction of the "new" atoms are old atoms that have been reexcited.) The motion of the new atoms has no correlation with that of the old, except that we may assume that the average excitation energy is about the same for the new atoms as for the old. When we add the x component of

radiation from all the atoms, we get the x component E_x of the total wave. It should have about the same amplitude as the E_x obtained from the old set of atoms. But the phase constant of the new E_x is completely unpredictable from the phase constant of the old E_x. The same statement holds for E_y. Furthermore, since the relative phase of the x and y motions of the new set of atoms does not have any correlation with the relative phase of the x and y motions for the old set, we see that the relative phase of E_x and E_y drifts in a completely unpredictable and "random" manner over time intervals long compared with τ.

We have assumed that the electron in the atom oscillates freely during the decay time τ. We have also assumed that the atom is at rest. Then the Fourier spectrum of the radiation emitted by a single atom has a bandwidth $\Delta\omega$ that is about equal to τ^{-1}. (Typical mean decay times are about 10^{-8} sec for atoms emitting visible light. That gives a bandwidth about 10^8 rad/sec.) In a gas discharge tube the atoms are not at rest but travel at velocities of order 10^5 cm/sec. This velocity gives a Doppler shift whose sign depends on whether the atom moves towards or away from the observer. The Doppler "broadening" gives a bandwidth that is greater than the "natural" width τ^{-1} by a factor of order 100. Furthermore, collisions often cut short the radiation process before the atom has had a chance to decay. In that case the bandwidth is further increased by "collision broadening."

When all factors of "frequency broadening" have been taken into account, we finally arrive at some bandwidth $\Delta\omega$, which may be much greater than τ^{-1}. In that case the time during which the polarization state may be taken to be roughly constant is not the natural decay time τ, but is what we shall call the "coherence time" t_{coh}, given by

$$t_{coh} \approx \frac{1}{\Delta\nu} \quad . \tag{31}$$

We can understand Eq. (31) as follows. The state of polarization remains essentially unchanged as long as the relative phase of E_x and E_y drifts by an amount that is small compared with 2π. Therefore the coherence time t_{coh} is given approximately by the time that it takes the maximum and minimum frequencies of the band to get out of phase by 2π:

$$\Delta\omega \, t_{coh} \approx 2\pi \ , \tag{32}$$

which is the same as Eq. (31).

The fact that there is a finite bandwidth $\Delta\omega$ does not necessarily mean that the polarization will change after a time interval of order $(\Delta\nu)^{-1}$. For example, there may be a piece of polaroid between the radiating atoms and the observer. In that case the x and y components of the radiation seen by the observer maintain a constant phase relation, although the bandwidth is still $\Delta\nu$. That is because the x and y components are no longer "independent." The polaroid has, so to speak, "examined" the x and y components of the incident radiation and at any instant has "selected for transmission" only those parts of the incident x and y components which superposed with the proper phase relation to drive the electrons in the polaroid along the easy direction, transverse to the "wires." Those parts of the incident x and y radiation which had a relative phase that drove electrons "along the wires" were absorbed.

Here is another example. Suppose we have two identical gas discharge light sources giving light of the same dominant frequency ω_o, same bandwidth, and same average intensity. With a suitable slab of glass or a "half-silvered"

mirror, we can make an arrangement such that the two sources appear to the observer to be superposed on top of one another (i.e., their images are superposed). Light from each source ends up going in the +z direction to the observer. Now we put a piece of polaroid over each source, so that one source gives radiation linearly polarized along $\hat{x}$, the other gives radiation linearly polarized along $\hat{y}$ (after both light beams are finally going in the +z direction). If the observer makes a measurement of polarization during a time interval short compared with the coherence time $(\triangle\nu)^{-1}$, he will find some definite polarization state. If he waits for a time interval long compared with $(\triangle\nu)^{-1}$ and makes another measurement of polarization, he will find a polarization state that is completely uncorrelated with the previous measurement. As a matter of fact, the observer will find it essentially impossible to distinguish this radiation from that which he would get by removing one of the sources and at the same time removing the piece of polaroid from the remaining source.

Unpolarized radiation. We are now prepared to say what is meant by "unpolarized" light. Unpolarized light is light whose two polarization components (x and y, or right-handed and left-handed) are emitted "independently" (i.e., are not locked in phase by a piece of polaroid, for example), and whose amplitudes and relative phase for the two polarization components have been measured by a technique that averages over a time interval long compared with the coherence time $(\triangle\nu)^{-1}$. There is no such thing as "intrinsically" unpolarized light. All you have to do to change "unpolarized" light into "completely polarized" light is to invent a technique that allows you to measure the polarization before the phases have a chance to drift.

Measurement of polarization. A quantitative description of the "amount of polarization", which means the amount of correlation between phases and amplitudes that is maintained during the measurement interval, can be given as follows: Suppose we represent the instantaneous state of polarization by A_1, A_2, ϕ_1 and ϕ_2 in the linear polarization representation, where $\underset{\sim}{E}$ is the real part of

$$\underset{\sim}{E}_c = e^{i(\omega_o t - k_o z)} \left[\hat{x}\, A_1\, e^{i\phi_1} + \hat{y}\, A_2\, e^{i\phi_2} \right] . \qquad (33)$$

We could write this as a superposition of various frequencies, but we prefer to think of Eq. (33) as a wave with dominant frequency ω_o and with amplitudes and phase constants A_1, A_2, ϕ_1, and ϕ_2 that are not quite constant in time but vary slowly (in some unpredictable manner).

Now let us see how to express A_1, A_2, ϕ_1 and ϕ_2 in terms of measurements of intensity only. (We use the word intensity as a synonym for energy flux.) That is the kind of measurement that is easiest to make. We imagine that we have polaroids and quarter-wave plates and also a photomultiplier with which we can measure the photon flux (number of photons per unit area per unit time) for any experimental set up. The average photon flux is proportional to the average classical energy flux. The average classical energy flux is proportional to the average over one cycle of the square of the electric field. The average over one cycle of the quantity $\cos^2\omega t$ is $\frac{1}{2}$. The average over one cycle of the quantity $|e^{i\omega t}|^2$ is 1 (since that is its constant value.) It is then easy to see that if we want to keep working with the complex expression given by Eq. (33), we can do so if we follow this rule: If

$$E(z,t) = \mathrm{Re}\, E_c(z,t) , \qquad (34)$$

then the time average of $E^2(z,t)$ over one period T, which we shall call $\langle E^2(z,T)\rangle$, is given by

$$\langle E^2(z,T)\rangle = \tfrac{1}{2} \left|E_c(z,t)\right|^2 . \qquad (35)$$

Here then is the procedure for finding the four constants of interest:

1. Rotate the axis (of easy transmission) of the polaroid to $\hat{x}$. Measure $\langle E_x{}^2(z,T)\rangle$, where T need not be as small as a single period provided that it is short compared with the coherence time $\Delta\nu^{-1}$. The result, according to Eqs. (33) and (35) is

$$\langle E_x{}^2 (z,T)\rangle = \tfrac{1}{2} A_1{}^2 , \qquad (36)$$

because the polaroid has absorbed the $\hat{y}$ component of $\underset{\sim}{E}$ and transmitted the $\hat{x}$ component undisturbed.

2. Rotate the polaroid to $\hat{y}$. Measure $\langle E_y{}^2(z,T)\rangle$, which gives

$$\langle E_y{}^2(z,T)\rangle = \tfrac{1}{2} A_2{}^2 . \qquad (37)$$

3. Rotate the polaroid to the 45° direction between $\hat{x}$ and $\hat{y}$. Call this direction $\hat{e}$. Then we have the unit vector $\hat{e}$ given by

$$\hat{e} = \frac{\hat{x} + \hat{y}}{\sqrt{2}} \qquad (38)$$

The electric field component transmitted by the polaroid is the dot product of $\hat{e}$ with the incident field $\underset{\sim}{E}$. Using the complex $\underset{\sim c}{E}$, we have

$$\hat{e} \cdot \underset{\sim c}{E}(z,t) = e^{i(\omega_o t - k_o z)} \left[\frac{A_1}{\sqrt{2}} e^{i\phi_1} + \frac{A_2}{\sqrt{2}} e^{i\phi_2} \right] \qquad (39)$$

The transmitted photon flux now gives us a measurement of

$$\langle(\hat{e}\cdot\underset{\sim}{E}(z,T)^2\rangle = \tfrac{1}{2}\left[\tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2^2 + A_1A_2\cos(\phi_1-\phi_2)\right] \qquad (40)$$

Since we have already determined A_1^2 and A_2^2 (A_1 and A_2 are positive real numbers) from Eqs. (36) and (37), we see that Eq. (40) gives us $\cos(\phi_1-\phi_2)$.

We still need $\sin(\phi_1-\phi_2)$ to pin down the relative phase. (We are not usually interested in the absolute phases.) To get it we use the quarter wave plate:

4. Leave the polaroid along $\hat{e}$ at 45° to $\hat{x}$ and $\hat{y}$. Then Eq. (39) still gives the transmitted field. Insert a quarter wave plate ahead of the polaroid with its slow axis along $\hat{x}$ or $\hat{y}$. For definiteness, suppose the slow axis is along $\hat{y}$. Then $\underset{\sim}{E}_c$ as given by Eq. (33) has ϕ_2 replaced by $\phi_2 - \frac{1}{2}\pi$. (Also, both ϕ_1 and ϕ_2 have acquired an uninteresting constant, which we drop.) Consequently, $\hat{e}\cdot\underset{\sim}{E}_c$ as given by Eq. (39) has ϕ_2 replaced by $\phi_2 - \frac{1}{2}\pi$. Now measure the photon flux behind the system of quarter wave plate plus polaroid. The flux is given by an expression like Eq. (40) with ϕ_2 replaced by $\phi_2 - \frac{1}{2}\pi$. Thus we determine

$$\langle(\hat{e}\cdot\underset{\sim}{E}(z,T))^2\rangle = \tfrac{1}{2}\left[\tfrac{1}{2}A_1^2 + \tfrac{1}{2}A_2^2 - A_1A_2\sin(\phi_1-\phi_2)\right] \qquad (41)$$

Thus we have completely determined A_1, A_2, and $\phi_1-\phi_2$ by the measurements represented by Eqs. (36), (37), (40) and (41). These are the results one gets if the measurement time is short compared with the coherence time.

Partial polarization. What will one get if the measurement time is not short compared with the coherence time? It depends on where the light came from, how it was prepared, etc. The light is said to be completely unpolarized if

all four of the intensity determinations represented by the left-hand sides of Eqs. (36), (37), (40), and (41) give exactly the same answer. Notice that this means that A_1 and A_2 are equal (on the average) and that both $\sin(\phi_1-\phi_2)$ and $\cos(\phi_1-\phi_2)$ average to zero. If not all of these conditions are met, the light is said to be partially polarized. It is possible to combine these four measurements into an expression which is zero for unpolarized light, unity for any pure polarization state, and something between zero and unity for intermediate cases of "partially polarized" light.

*Time coherence and optical beats. One can obtain interference between waves of different frequency. That holds with optical phenomena as well as other phenomena. Suppose we have two light waves 1 and 2 producing electric fields $\underset{\sim}{E}_1$ and $\underset{\sim}{E}_2$, both polarized (let us say) along $\hat{x}$. The total field at fixed z is the superposition of $\underset{\sim}{E}_1$ and $\underset{\sim}{E}_2$. Using a complex field $E_c(t)$ we have

$$E_c(t) = E_1 e^{i\omega_1 t} e^{i\phi_1} + E_2 e^{i\omega_2 t} e^{i\phi_2} . \qquad (42)$$

The energy flux, which can be measured by a photomultiplier (whose output current is proportional to the incident energy flux), is proportional to the average of $E^2(t)$ over one period:

$$\langle E^2(T)\rangle = \tfrac{1}{2}\,|E_c(t)|^2 = \tfrac{1}{2}\left\{E_1^{\,2} + E_2^{\,2} + 2E_1E_2\cos\left[(\omega_1-\omega_2)t + (\phi_1-\phi_2)\right]\right\} \qquad (43)$$

Thus one may hope to measure a photomultiplier current that varies at the beat frequency $\omega_1-\omega_2$. What are the requirements on bandwidth? Remember that our simple point of view is to think of the amplitudes and phase constants as slowly varying in an unpredictable manner, with (for example) ϕ_1 drifting randomly by an amount of order 2π in a coherence time interval which is the inverse of the bandwidth of oscillation 1:

$$t_1(\text{coherence}) \approx (\Delta\nu_1)^{-1} , \tag{44}$$

$$t_2(\text{coherence}) \approx (\Delta\nu_2)^{-1} . \tag{45}$$

It is clear that if we are to observe beats, the individual components must keep their phases roughly constant for the duration of a beat period. Thus to observe beats we require that both coherence times be long compared with the beat period, i.e., that both bandwidths be small compared with the beat frequency:

$$\Delta\nu_1 \text{ and } \Delta\nu_2 \text{ both} < |\nu_1-\nu_2| \text{ for observable beats.} \tag{47}$$

In addition you need to be able to detect variations in the photomultiplier current at the beat frequency. It also helps to be ingenious. The experiment has been done, and it is a beauty.*

*Photoelectric Mixing of Incoherent Light by A.T. Forrester, R.A. Gudmundsen, and P.O. Johnson, Phys. Rev. 99, 169 (1955).

Problems and Home Experiments - Chapter 4.

1. Carry out explicitly the steps outlined following Eq. (11) to show that Eq. (11) represents a displacement $\underset{\sim}{\Psi}(t)$ that follows an elliptical path.

2. Show that the amplitudes and phase constants for the linear polarization representation of Eq. (11) are related to those for the circular polarization representation of Eq. (12) by (taking all amplitudes A to be positive)

$$A_+^2 = \tfrac{1}{4}\left[A_1^2 + A_2^2 + 2A_1A_2 \sin(\phi_1-\phi_2)\right]$$
$$A_-^2 = \tfrac{1}{4}\left[A_1^2 + A_2^2 - 2A_1A_2 \sin(\phi_1-\phi_2)\right] \qquad \text{(P1)}$$

$$A_+ \cos\phi_+ = \tfrac{1}{2}[A_1 \cos\phi_1 - A_2 \sin\phi_2]$$
$$A_+ \sin\phi_+ = \tfrac{1}{2}[A_1 \sin\phi_1 + A_2 \cos\phi_2]$$
$$A_- \cos\phi_- = \tfrac{1}{2}[A_1 \cos\phi_1 + A_2 \sin\phi_2]$$
$$A_- \sin\phi_- = \tfrac{1}{2}[A_1 \sin\phi_1 - A_2 \cos\phi_2]$$

Hint: Use complex numbers to do away with most of the algebra.

3. Circularly polarized light of intensity I_o (intensity means energy flux per unit area per unit time; this is proportional to a photomultiplier's output current, at a given frequency) is incident on a single polaroid. Show that the output intensity (intensity of the light emerging from the rear of the polaroid) is $\frac{1}{2} I_o$.

4. Linearly polarized light with polarization direction at angle θ from $\hat{x}$ is incident on a polaroid with easy axis along $\hat{x}$. The first polaroid is followed by a second polaroid with its easy axis along the direction of polarization of the original incident light. Show that if the input intensity is I_o, the output intensity is $\cos^4\theta\, I_o$.

5. Circularly polarized light of intensity I_o is incident on a sandwich of three polaroids. The first and third polaroids are crossed, i.e., their easy axes are at 90° to one another. The middle polaroid makes an angle θ with the axis of the first polaroid. Show that the output intensity is $\frac{1}{2} I_o \cos^2 \theta \sin^2 \theta$.

6. A very large number N + 1 of polaroids are arranged in a sandwich. The angle of the easy axis of each polaroid is a constant angle α greater than that of its immediate predecessor. Thus the last polaroid is at an angle $\theta = N\alpha$ from the first. Neglecting any losses due to reflection at the many surfaces, and supposing that linearly polarized light of intensity I_o is incident on the first polaroid with its polarization along the easy axis of the first polaroid, find the output intensity. Take N to be very large and retain only the first interesting terms in an appropriate power series (Taylor's expansion). Answer:

$$I = I_o \left(1 - \frac{\theta^2}{N} + \text{higher order terms}\right). \qquad (P2)$$

That means that even if θ is 90°, so that the first and last polaroids are "crossed", the output intensity is equal to the input intensity if we have enough intermediate polaroids. We can "gently" rotate the plane of polarization and lose nothing! Thus if we cemented a large number of polaroids together (using clear cement with the same index as the polaroids, so as to minimize reflections), we would have something like a gigantic sugar molecule, which rotates the plane of polarization without absorbing any energy.

Another way to get "macroscopic optical activity" is to twist pieces of tinfoil into corkscrews (all with the same handedness), embed them in styrofoam (a good approximation to a massless, rigid, electron-free supporting medium), and send linearly polarized microwaves through the stuff. The plane

of polarization is rotated.

7. Suppose you have linearly polarized incident light with polarization along $\hat{x}$. You desire linearly polarized light with polarization at 30° to $\hat{x}$, i.e., along

$$\hat{e} = \hat{x} \cos 30^\circ + \hat{y} \sin 30^\circ.$$

How can you obtain this transmitted field (a) at the cost of some loss of intensity, (b) without loss of intensity and without using any polaroids?

8. What is the transmitted intensity for unpolarized light of intensity I_o incident on crossed polaroids with a half-wave plate between them (a) when the retardation plate's optic axis (say the slow axis) is parallel to the easy axis of one of the polaroids; (b) when the wave plate's optic axis is at 45° to one of the easy axes?

9. Answer the same questions as in problem 8, but use a quarter-wave plate.

10. Home Experiment. Test various sheets of plastic (drawing instruments, cellophane, scotch tape, etc.) for double refraction by rotating them between crossed polaroids. How can you tell if you are lucky enough to find a quarter-wave plate or half-wave plate? Try the experiment of stretching Saranwrap to make it doubly refracting. Perhaps by dipping it in iodine while stretching it, you can make your own polaroid (the author has not tried this yet). Try adding successive layers of scotch cellophane tape on top of one another on a microscope slide to see if you can make a good quarter- or half-wave plate. (Microscope slides are about 25 ¢/dozen. Get some.) In order to understand what you are doing, you should use a red or blue cellophane filter, because the double refraction is very dependent on wave length. That is not so much because the indices n_s and n_f vary with wavelength as it is because the phase difference, given by

$$\Delta\phi = (n_s - n_f)\, 2\pi \frac{\Delta z}{\lambda_{vac}} \quad , \qquad (P3)$$

has that factor λ_{vac} in the denominator. For blue light, λ_{vac} is about 4500 A°. For red, λ_{vac} is about 6500 A°. Green is in the middle of the visible spectrum at about 5500 A°. Thus over the visible range, λ_{vac} varies by a factor of about $6500/4500 \approx 1.4$. Thus a sheet that is a quarter-wave plate for red light will be a 1.4-quarter-wave plate, i.e., a 0.35-wave plate for blue. Actually you will find a much larger variation than that, because what you have if you luckily find a "quarter-wave plate" is probably not one quarter, but a large number of full cycles plus one odd quarter cycle. Thus if a piece of scotch tape that behaves like a quarter-wave plate for red light is actually a 17-quarter-wave plate, then for blue light it will be a $1.4 \times 17 = 24.5$ quarter = 6 full waves plus one eighth wave. The effect of varying wavelength is thus magnified by the number of "useless" full cycles of relative phase shift. The way to reduce the color dependence is to make a quarter-wave plate truly a <u>one</u> quarter and not a $(4N + 1)$-quarter-wave plate. That is how Polaroid makes their plates fairly "broad-band" (i.e., so that they have about the same phase shift over the entire visible spectrum).

11. <u>Home Experiment. Determination of $n_s - n_f$ for scotch tape.</u> (We mean the clear cellophane tape, not the translucent "magic mending tape.") This experiment was stimulated by Home Experiment 10. (Read that first.) Here is the experiment. Put about 15 layers of scotch tape on a microscope slide. Make the following sandwich: polaroid number 1 at 45° to the horizontal; diffraction grating oriented so that it spreads out the colors of a light bulb horizontally; microscopic slide (with scotch tape) horizontal; second polaroid with easy axis parallel to the first. Let the slide cover only half of the polaroids, so that you can look at the spectrum of a light bulb (the first order spectrum

of the grating) either through the parallel polaroids alone or through the scotch tape between the parallel polaroids. You should see several "black" bands in the color spectrum of the light bulb when you look through the scotch tape. The black bands are separated by 2π in relative phase shift of the slow and fast components; that is the heart of the experiment. If you know the wave lengths at two neighboring black bands you can use Eq. (P3) to find $(n_s-n_f)\Delta z$. The easiest way to estimate the wavelengths at the black bands is to move your head up and down so that you can look back and forth between the unbanded full spectrum and the banded one. Then you can tell what colors are absorbed by the second polaroid. Then find a good color plate that shows the continuous spectrum and is labeled with wavelengths. (This book will have one, we hope.)

Now that you know $(n_s-n_f)\ \Delta z$ for 15 layers (if that is the number you used), you can divide by 15 to find it for one layer. Then use Eq. (P3) to find the phase shift for, say, green light of wavelength 5500 A^o. Answer: for my roll of tape and green light, I get that one layer is about one-half-wave plate (with no extra full cycles). (To check, put a slide with one layer of tape at 45° between crossed polaroids. Is the transmitted light green? Should it be?) Now measure the thickness of the 15 layers, either by using a micrometer or by reading the label on the box of tape to find out how many inches of tape there are and then doing some geometry about the inner and outer radii of the spool of tape. My tape has thickness $\approx$ 1.8 mils (thousandths of an inch) and thus I find $n_s-n_f \approx 0.040$.

The average index of refraction can be measured by some cruder technique, for example by measuring Brewster's angle, as in Home Experiment 19. The answer is that the average index is about 1.5.

12. Fractional polarization of blue sky. Look at Fig. 6 and its caption. Define the fractional polarization P by

$$P = \frac{I(\hat{x}_1) - I(\hat{y}_1)}{I(\hat{x}_1) + I(\hat{y}_1)} \quad , \tag{P4}$$

where $I(\hat{x}_1)$ is the intensity of scattered light polarized along $\hat{x}_1$ and $I(\hat{y}_1)$ is the intensity of scattered light polarized along the projection of $\hat{y}_1$ seen by the observer. Show that P has a dependence on the scattering angle θ_{scatt} of Fig. 6 given by

$$P = (1 - \cos^2\theta_{scatt})/(1 + \cos^2\theta_{scatt}) \; . \tag{P5}$$

Notice that P is zero at scattering of 0° or 180° and is 1 at 90°.

13. Optical activity. Suppose you send linearly polarized light through a length L of Karo corn syrup and find that, for L = 5 cm, red light is rotated by 45°. Now reflect the light that has passed through the syrup from a mirror and send it back through the syrup, so that the total length is 10 cm. (If you do the experiment, make the angle of reflection not quite 180°; then look at the "image light bulb" through both the "real syrup" and the "image syrup." As a control experiment, you can look through the image syrup alone by moving your head.) Question: after the two traversals, is the linear polarization at 0° or 90° to the original direction? Second question: Suppose you use two mirrors to reverse the direction of the light beam. The first mirror is set at 45° to the beam and deflects the beam by 90°. The second mirror is set at 45° to the once-deflected beam and deflects it by another 90°, so that it has finally been deflected through 180°. It now goes through the sugar on its return path. Is the polarization at 0° or 90° to the original direction?

14. Suppose that a beam of linearly polarized light is incident on a half-wave plate which is rotating about the beam axis with angular velocity ω_o.

Show that the output light is linearly polarized with the polarization direction rotating at $2\omega_o$.

15. Home Experiment. Is the white light scattered from clouds polarized? Do the experiment with your polaroid. Explain the result.

16. Home Experiment. Is the rainbow polarized? Predict the result, then do the experiment. (A garden hose set for fine spray can replace the rain. A streetlight can replace the sun.)

17. Home Experiment. When the moon appears "half-full," the illuminated portion is scattering sunlight through about 90° to your eye. We know that for 90° scattering the blue sky is almost completely linearly polarized. Do you predict that half-moon light is polarized? Do the experiment. Now think about how the earth looks from the moon at "half-earth." Is the earthlight polarized? You can look for twenty-four hours while the earth turns. Answer: Sometimes; it depends on the time and the weather. Why?

18. Home Experiment. Look at a light bulb through a piece of polaroid. Is the light polarized? Now insert a piece of cellophane (or your quarter- or half-wave plate) between the light bulb and the polaroid. Now is the light polarized? Reflect the light from a silvery metal, like a table knife. Is the reflected light polarized?

19. Home Experiment. Measuring the index of refraction by finding Brewster's angle. You need a light bulb (perhaps covered with a piece of cardboard with a hole to get a reasonably small source), a piece of glass, a table, a cardboard box or something to give a measurable location for your eye, and a single piece of polaroid. Lay the piece of glass flat on the table and look at the reflection of the bulb. (You will see two reflections, one from the front and one from the

rear surface. If you wish to, you can eliminate the one from the rear surface by spraying the rear surface of the glass with black paint.) Vary the angles until the polaroid reveals that the reflected light is completely polarized. Measure the appropriate distances and obtain the index of refraction by the formula for Brewster's angle, $\tan \theta_B = n$. With this crude set-up you cannot measure better than to a few degrees, so that you probably cannot distinguish Brewster's angle for glass from that for a smooth surface of water.

20. Home Experiment. Phase relations in specular reflection of light from glass. We are trying to verify the relationships shown in Fig. 8. The experimental setup is as in Home Experiment 19, except that between the piece of glass lying on the table and the light we place a piece of polaroid with easy axis rotated at 45° to the horizontal. (Suggestion: For a convenient mounting, put a little non-hardening putty or glazing compound on a microscope slide and stick the corner of the polaroid into the putty. The microscope slide can be the glass surface used for reflecting the light.)

Suppose that when your eye is at the microscope slide and you are looking back at the light bulb through the first polaroid, the easy direction of the polaroid is "upper right to lower left." Keep analyzing the polarization of the reflected light as you vary the angle of incidence by changing the position of the slide or light bulb. You will find that near normal incidence the polarization of the reflected light is along "upper left to lower right." As you move the slide and approach Brewster's angle, the polarization remains linear, but it rotates toward the horizontal. It becomes horizontal at Brewster's angle and then continues to rotate in the same direction as you go beyond Brewster's angle towards grazing incidence; i.e., it becomes "lower left to upper right." Thus in going from normal to grazing incidence the polarization

rotates by 90°, as predicted by Fig. 8. (At normal incidence both components are equally well reflected, as they must be, since they don't know which is which, so to speak. Therefore the polarization is at 45°. At grazing incidence both components are almost completely reflected, so that they are equally well reflected. Therefore the polarization is at 45°.) It is interesting to notice that the polarization remains linear through all of this experiment. This means that there are no phase shifts other than 0° or 180° introduced between those components in the plane of incidence and those perpendicular to it. Thus the incident waves always experience a purely resistive load when they reflect. That is what we expect for reflection from a transparent medium.

21. Prove the statement made in numbered paragraph 8 following Eq. (29), where it is stated that a right-hand circular polarizer will completely pass incident right-handed circularly polarized light traveling in the reverse direction (i.e., incident on the quarter-wave-plate face) but will completely absorb left-handed light traveling in the reverse direction. A right-hand circular polarizer followed by a half-wave plate is a left-hand circular polarizer. It transmits left-handed light in the reverse direction and absorbs right-handed light in the reverse direction.

It should be apparent that right- and left-handed circular polarizers are useful in analysing the polarization of light.

22. Home Experiment. Take your circular polarizer and lay it on top of a shiny piece of metal, like a dime or a tableknife. First lay it with one face up, then with the other. Explain what you see! Hint: see problem 21. Also remember that reflection changes a left hand into a right hand.

23. Home Experiment. Lay the circular polarizer on a dime (or other mirror-like surface) with the quarter-wave-plate face on the under side, so that the dime looks "black" (or actually, dark blue). Then slip your quarter-wave

plate under the circular polarizer. Explain what you see. Now do the same with the half-wave plate. Explain.

24. Home Experiment. If you lay your circular polarizer on a sheet of white paper, you do not get the dramatic darkening effect that you observed with the shiny dime. Why not?

25. Home Experiment. Instead of a dime, use a smooth piece of aluminum foil in Home Experiment 22. (Don't crinkle it; keep it glossy smooth.) Now put a vee-shaped crease in the foil, so that the foil is like a smooth plain with a vee-shaped valley running along it. Lay the circular polarizer on the foil as before. The crease is not dark, it is bright. Why? Now crinkle the foil all up into a grainy rough mess and repeat the experiment. Explain your results.

26. Home Experiment. Phase changes in metallic reflection. The experimental set-up is like that for Home Experiment 20. But instead of a piece of glass use any shiny flat piece of metal, for example the stainless steel blade of a table or kitchen knife, or any chromium or silver-plated object. (Do not use an ordinary mirror, i.e., a rear-surfaced piece of glass. It won't work.) You will need two polaroids and one quarter-wave-plate. First, verify that $\parallel$ (parallel) or $\perp$ (perpendicular) linearly-polarized light (i.e., respectively parallel or perpendicular to the plane of incidence) retains its polarization upon reflection. (That is analogous to the action of a retardation plate on linearly polarized light polarized parallel or perpendicular to the optic axis; nothing happens to the polarization.) Next, linearly polarize the incident light at 45° to the plane of incidence. Set the angle of incidence so that if the light bulb is 1 ft above the table, the knife is about 3 ft from the light bulb. Now analyse the reflected light using your polaroid and quarter-wave plate (or using your circular polarizer "run backwards" as an analyser, with or without

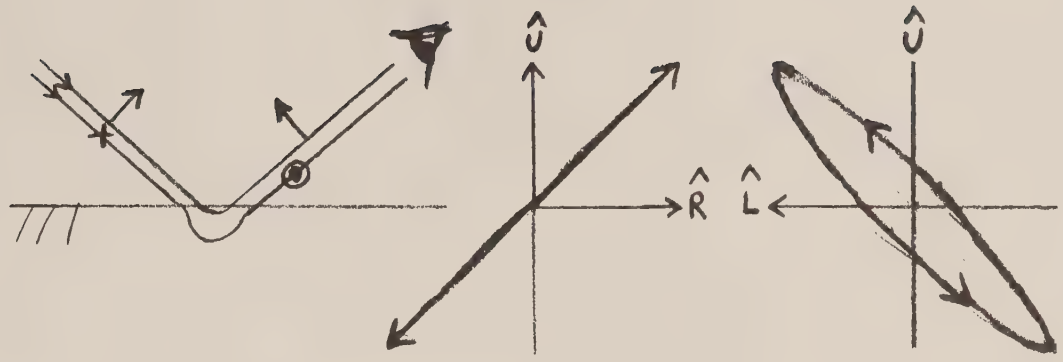
a half-wave plate, as discussed in problem 21). You will find that the polarization is elliptical. By varying the angle of incidence, you can find a spot where the polarization of the reflected light is almost perfectly circular. If you now tilt the polarizing polaroid by five or ten degrees away from 45°, tilting the easy axis towards the vertical so as to increase the $\|$ component slightly, you can get completely circularly polarized reflected light. (This slight tilt is necessary to compensate for the fact that the $\|$ component is not as completely reflected as is the $\perp$ component.)

By rotating the polarizing polaroid from an "upper right" to an "upper left" direction of the easy transmission axis, the handedness of the reflected light is reversed.

Here is a qualitative explanation of the phenomenon. The metal is a reactive medium. Both polarization components of the incident light are almost totally reflected. There is a phase shift corresponding to the time for the fields to penetrate into the reactive medium for about one exponential attenuation distance (on the average) and then turn around and come back out. The $\|$ and $\perp$ polarization components do not have the same phase shift for the following reason. The $\perp$ component is parallel to the surface. The electrons are free to move parallel to the surface, and they move so as to cancel the incident radiation. The time delay and phase shift is due to the inertia of the electrons. The $\perp$ component therefore has a certain phase retardation due to this time delay. Now consider the $\|$ component. At near normal incidence the $\|$ component is nearly parallel to the surface and therefore behaves like the $\perp$ component. The phase retardation is therefore the same as for the $\perp$ component. Both components receive a minus sign upon reflection, in addition to the small phase shift due to penetration. Therefore the incident polarization "upper right" (as seen looking back at the light bulb from the metallic reflector) becomes "upper left" after reflection.

(Mnemonic: upper is still upper after reflection, i.e., your head is still above your feet when you look in a mirror. But left becomes right.) But now suppose we are not near normal incidence. Then the $\|$ component of electric field is not parallel to the surface. We can resolve it into a component parallel to the surface, and a component perpendicular to the surface. That component parallel to the surface continues to behave in the usual way and undergoes the same phase retardation as before. But the component perpendicular to the surface behaves in a completely different manner. Charges are not free to move perpendicular to the surface. The surface acquires a surface charge and the charges come to rest very rapidly. The time delay due to the inertia of the electrons moving parallel to the surface is not present for the motion perpendicular to the surface, because the motion is so tiny. Therefore this part of the $\|$ component reflects with negligible time delay. Also, it does not acquire a minus sign upon reflection. (Recall that in electrostatics an electric field perpendicular to a surface of a conductor is not cancelled to zero outside the conductor by the fields set up by the charges. A surface charge is induced that cancels the field inside the conductor but not outside.)

The net result is that an incident linear polarization of "upper right" becomes, to first approximation, a linear polarization of "upper left." But (in the next approximation) the "upper part reflects with less time delay than the "left" part. Thus the "upper" leads the "left" slightly in phase, and we see that the sense of rotation of the resulting elliptically polarized light is counterclockwise (for our example). An easy way to remember it is to follow the direction of rotation which takes the original linear polarization direction from 45° to the vertical through the vertical to 45° on the other side of the vertical; that is the sense of rotation the elliptically polarized light acquires. This is shown schematically in the accompanying figure:



(a) picture (b) incident polarization, "upper-right" (c) reflected polarization, left-handed

Fig. P1. Metallic reflection. In part (a) the cross and dot represent electric field into and out of the paper, respectively. The diagrams in parts (b) and (c) correspond to the directions of fields shown in part (a). Part (c) also shows the effect of the greater phase retardation experienced by the ⊥ component. This is indicated in part (a) by a deeper "penetration" of the ⊥ component into the metal. In parts (b) and (c) the unit vectors $\hat{U}$, $\hat{R}$, and $\hat{L}$ stand for "up," "right," and "left," respectively.

The relative phase shift increases monotonically as the angle of incidence increases. At (typically) about 75° angle of incidence, the relative phase shift reaches 90°, and one gets nearly circularly polarized reflected light. (It would be exactly circularly polarized if the ⊥ and ∥ components were both completely reflected.)

If you try this experiment with an ordinary mirror consisting of a rear silvered piece of glass it won't work very well. Reflection from the front

glass surface is superposed on that from the silvered surface. But the glass is not perfectly flat and uniform in thickness. Therefore when you look at a light bulb of any reasonable size--a few millimeters, say--the phase of the superposition from the two surfaces varies a lot over this region, and on the average the polarization is washed out. At near-normal incidence the polarization does not wash out, because the metal and the glass both give essentially the same thing--linear polarization at 90° to the incident polarization (this is exactly true for the glass and approximately true for the metal--the ellipse is almost a line).

If you try this experiment with a smooth metal surface that has been coated with a thin dielectric film, you may be shocked at the result. Interference between the two surfaces is the same over a broad region (if the film is a uniform one), and one gets a "coherent" superposition of the linear polarization from the top surface (the dielectric) and the elliptical polarization from the metal. The effect can be so drastic that it reverses the sense of rotation of the resultant reflected light. Therefore, if you happen to have a front-surfaced mirror with a "protective coating" and wish to use it in the present Home Experiment, beware!

27. Home Experiment. Determining the handedness of your circular polarizer. Perhaps the label is wrong. How can you find out if your circular polarizer really makes left-handed and not right-handed circularly polarized light ? Here is one way. Learn how to use metallic reflection to get circularly polarized light, as in Home Experiment 26. Accept the theory, which says that the sense of rotation is as shown in Fig. P1. Now use your circular polarizer "run backwards" as an analyser. If you have produced left-handed circularly polarized light with the metal, and if your polarizer is a left-handed one, it will pass the light from the metal, and it will absorb it if you add a half-wave plate in

front of your polarizer. [Note: either my circular polarizer is labeled incorrectly (or with a different convention for handedness), or else Fig. P1 is wrong. I believe the figure.]

28. Home Experiment. Effective spring constants for Saranwrap molecules. Stretch a piece of Saranwrap and lay it behind a polaroid at 45° to the easy axis of the polaroid. Don't stretch too hard--we don't want to go beyond a relative phase shift of $\frac{1}{2}\pi$. Now determine the handedness of the elliptically polarized light. You can do this with your circular polarizer which you calibrated for handedness in Home Experiment 27. Once you know the handedness, you know whether the axis of stretch is the slow or fast axis. Now assume that the stretching lines up the molecules with their long dimension along the stretching direction. You now know whether the index of refraction is larger or smaller along the long direction of the molecule. Large index of refraction means large dielectric constant, which means large molecular polarizability, which means weak effective spring constant (as long as the frequency of the light is less than the effective natural vibration frequency of the electrons in the molecule. That is the case for visible light in glass, as we found in Problem 40, Chapter 3. We can assume it is also the case here.) Thus if the stretch axis turns out to be the slow axis, it means that the effective spring constant for vibration along the molecule is less than for vibration perpendicular to it--provided all of our assumptions also hold.

29. Home Experiment. Is the lengthwise direction along a piece of scotch tape the slow axis or the fast axis? See Home Experiment 28 for the experimental technique.

Chapter 5 - Interference and Diffraction

Section		Page
5.1	Introduction.	1
5.2	Interference between two coherent point sources.	2
5.3	Interference between two independent sources.	13
5.4	How large can a "point" light source be?	20
5.5	Angular width of a "beam" of traveling waves -- approximate treatment.	24
5.6	Multiple-slit interference pattern and single-slit diffraction pattern.	31
5.7	Geometrical optics	45
	Problems and Home Experiments	62

5.1 Introduction

Most of our studies so far have been essentially one-dimensional in the sense that there was only one path by which a wave emitted at one place could go to another place. Now we shall consider situations where there are different possible paths from an emitter to a detector. These lead to what are called interference or diffraction phenomena, resulting from constructive and destructive superposition of waves that have different phase shifts depending on the path taken.

In Sec. 5.2 we consider the superposition at a detector of the waves emitted by two point sources having the same frequency and a constant phase relation. Examples are water waves emitted by two screw heads jiggling the surface of a pan of water (Home Experiment 21) or light emitted by the currents in the edges of two slits which are illuminated by a line or point source (Home Experiment 16).

In Sec. 5.3 we consider interference between two "independent" sources, i.e., ones whose phases are not constrained to maintain a definite relation. We find that the interference pattern remains constant only for time intervals of order $(\Delta\nu)^{-1}$, where $\Delta\nu$ is the frequency bandwith of the sources. By a sufficiently fast measurement one can nevertheless determine the interference pattern.

In Sec. 5.4 we find how large a source can be and still behave like a point source, when the source consists of independently radiating parts, and when the detector averages over long time intervals [i.e., long compared with $(\Delta\nu)^{-1}$]. The result can be verified in an easy home experiment (Home Experiment 17). Another home experiment (Home Experiment 18) demonstrates the coherence of a Lloyd's mirror.

In Sec. 5.5 we give a crude derivation of the result that a beam of width D has an angular divergence of order $\Delta\theta \approx \lambda/D$ about the dominant direction of travel. This fact is mathematically related to the fact that a pulse of time width Δt has a frequency width of order $(\Delta t)^{-1}$. From the point of view of the mathematical theory of Fourier analysis, these two relationships are the same.

In Sec. 5.6 we consider interference due to many slits and also that due to a wide single slit. The emphasis is on optical and electromagnetic phenomena. Several home experiments involve diffraction gratings and various diffraction patterns. For these experiments we strongly advise the student to get a "display lamp" -- a light bulb with a clear glass envelope and a single straight filament about 3 inches long (about 40 cents in most grocery or hardware stores). Most of the experiments use one of these as a line source.

In Sec. 5.7 we study so-called "geometric" optics. We first derive the law of specular reflection and Snell's law of refraction from the wave properties of light. Then we consider various mirrors, prisms, and thin lenses.

5.2 Interference between two coherent point sources.

The simplest situation involving interference is that in which there are two identical point sources at different locations, each emitting harmonic traveling waves of the same frequency into an open homogeneous medium. If each source has a definite frequency (rather than a dominant frequency and a finite frequency bandwidth), then the relative phase of the two sources (the difference between their phase constants) does not change with time. The sources are then said to be coherent. (Even if they have

different frequencies, they are "coherent" if each is monochromatic, since their relative phase is always completely determined.) If each source has the same dominant frequency and each has a finite bandwidth $\Delta\nu$, then, if the sources are "independent," the relative phase of the two sources will only remain constant over times of the order of $(\Delta\nu)^{-1}$. On the other hand, if the two sources are not independent but are "locked" in phase with one another, then even though the phase constant of each source will drift in an uncontrollable manner through a phase of order 2π in a time $(\Delta\nu)^{-1}$, the relative phase will remain constant. The sources are then said to be coherent even though they are not monochromatic.

As an example of two coherent sources of waves, consider two rods which touch the surface of a body of water. If the rods are identically driven in vertical oscillations, they produce surface-tension waves on the water. The relative phase of the rods is constant because they are driven by a common source. As another example of two coherent sources, consider two identical radio antennas driven at constant relative phase by the same oscillator. Even if the oscillator is not perfectly monochromatic, the relative phase of the two antenna currents remains constant. As an example of two coherent sources of visible light, consider two small holes or parallel slits in an opaque screen which is illuminated on one side by a distant "point" source of light. Currents are induced in the edges of the slits by the electric field of the electromagnetic radiation ("light") emitted by the point source. The two slits are then said to be coherent sources of light. See Figure 1.

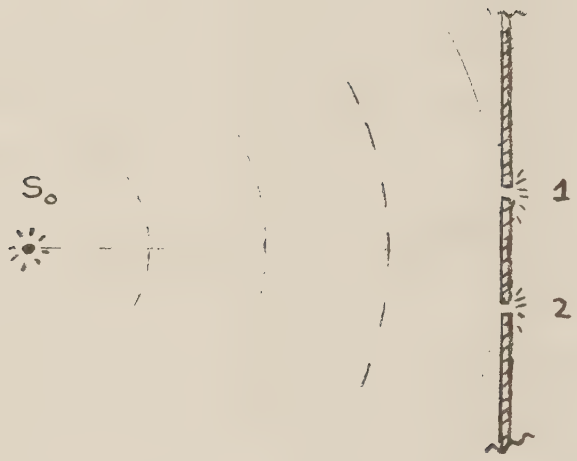


Fig. 5-1. Two coherent sources of light. Currents in the edges of slits 1 and 2 are driven by incident waves emitted by point source S_o. The phase constant of S_o may drift or change suddenly, but the relative phase of the slit currents remains constant.

In all of these examples we need a "detector" that is responsive to the waves. In the case of the surface-tension waves, we may use a tiny piece of cork which floats on the surface and whose vertical displacement can be measured. In the case of the radio waves we may use a detector consisting of a receiving antenna, a tuned resonant circuit, and a source of power to replenish the energy losses of the tuned circuit. In the case of the visible light we may use our eye, or a photographic emulsion, or we may use a photomultiplier whose output current we can measure. In any case, the detector will experience a total wave that is the linear superposition of two contributions, one from each source.

For some locations of the detector, the arrival of a wave crest (or trough) from one source is always accompanied by the simultaneous arrival

of a crest (or trough) from the other source. Such a location is called a region of constructive interference. At other locations the arrival of a crest from one source is always accompanied by the arrival of a trough from the other, and we then have a region of destructive interference. Since(by hypothesis) the two sources maintain a constant relative phase, a region that is one of constructive interference at a given time will always be a region of constructive interference, and likewise a region of destructive interference at a given time remains one for all time. The pattern formed by the various regions of interference maxima and minima is called an interference pattern. Even though the waves are traveling waves, the interference pattern is stationary in the sense described above.

Notice that even if the oscillator that drives the two antennas is turned off and then on again with a new phase constant, the relative phase of the antenna currents remains unchanged. Similarly if the point source driving the two slits is turned off and on, the slit currents maintain constant relative phase. Therefore the interference pattern is unchanged. On the other hand, if the point source is moved so as to change the distance to one slit by a different amount from that to the other slit, the relative phase of the induced currents will change, and the locations of interference maxima and minima will change, i.e., the interference pattern will change. Similarly if we insert a delay cable between the oscillator and one of the antennas so as to change the relative phase of the antenna currents, that will change the interference pattern.

In all the cases which interest us, the detector is at a distance from the two sources which is large compared with the separation of the sources. Therefore, as far as the effect of distance on wave amplitude is concerned,

we can say that the two identical sources are essentially at the same distance from the detector, so that each source (if present by itself) will contribute a traveling wave having the same amplitude as that contributed by the other. The time dependence of the total wave function at a given position is therefore given by superposition of two harmonic oscillations having the same frequency and amplitude but having (generally) different phase constants. The two phase constants (at a given detector location) depend on the phase constants of the two oscillating sources and on the number of wavelengths between each source and the field point. (The location of the detector is often called the field point, P .) If the distance from the field point P to one source equals that to the other source or they differ by a whole number of wavelengths, and if the sources oscillate in phase, then P is at an interference maximum and the amplitude of its harmonic oscillation is twice the amplitude it would have if either source were present alone. (If the sources oscillate 180° our of phase, P is at an interference node and has zero amplitude.) If the distance from the field point P to one source exceeds that to the other by $\frac{1}{2}\lambda$ (plus any whole number of wavelengths) and the sources oscillate in phase, then P is at an interference node and has zero amplitude.

Geometrical relations for distant field point. In Fig. 2 we show two point sources emitting waves that are detected at a distant field point P .

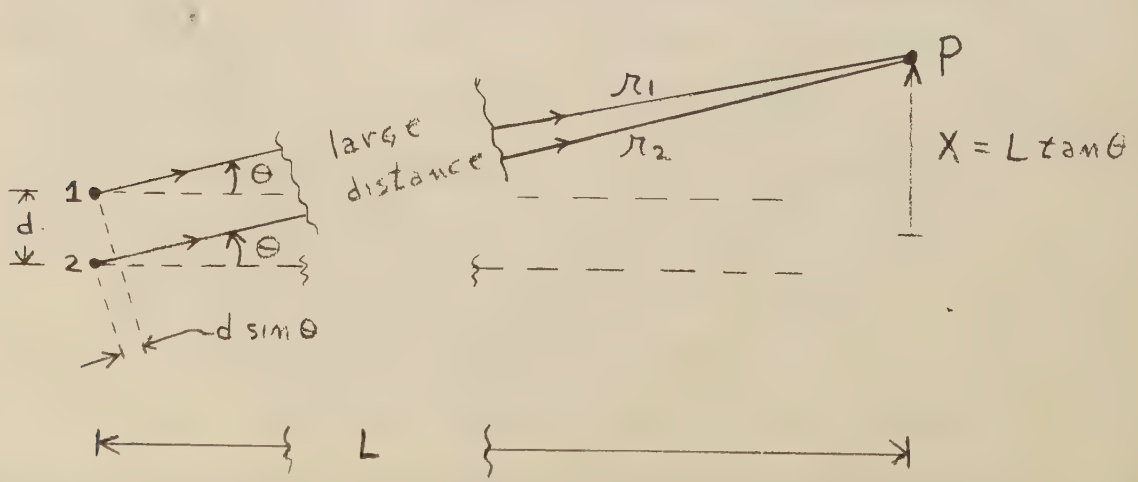


Fig. 5-2 Two point sources emitting waves which are detected at a distant field point P .

We are only going to look at the interference pattern in the plane containing the two sources and the field point P. Our results will also apply to two "line" sources (consisting of slits in the case of light), or to two radio antennas oriented perpendicular to the plane of the figure, or to surface waves in water.

When the distances r_1 and r_2 from sources 1 and 2 to the field point P are large compared with the separation d, then the two "rays" along the lines of sight from the two sources to point P are nearly parallel, both being at the same angle θ to the z axis as shown in the figure. In that case the path difference $r_2 - r_1$ is essentially equal to $d \sin \theta$. Therefore if the two sources oscillate in phase, P lies at a region of constructive interference when $d \sin \theta = 0, \pm \lambda, \pm 2\lambda$, etc. The interference maximum at $\theta = 0$ is called the principal or zeroth order maximum. The first maximum on either side, where $d \sin \theta$ is $\pm \lambda$, is called the first order maximum, etc. The regions of destructive interference, where the total wave is always zero, are called nodes. They occur at angles where the path difference $d \sin \theta$ is $\pm \frac{1}{2}\lambda, \pm 3/2\,\lambda$, etc.

Total wave for sources oscillating in phase. We now derive an expression for the total amplitude under the assumption that both sources oscillate in phase. The wave from source 1 has the form $A \cos(\omega t - kr_1)$; that from source 2 has the form $A \cos(\omega t - kr_2)$. We have taken both the amplitudes to be equal to A and both the phase constants to be equal to zero. The amplitude A depends on the distance from the sources to P, so that we can write $A(r)$ instead of A, with r denoting the average of r_1 and r_2. The total wave is thus given by the real part of the complex wave function $\psi_c(r,\theta,t)$:

$$\psi_c(r,\theta,t) = A(r)e^{i\omega t}\, e^{-ikr_1} + A(r)e^{i\omega t}\, e^{-ikr_2} \quad . \qquad (1)$$

If we define r as the average of r_1 and r_2

$$r = \tfrac{1}{2}(r_1 + r_2) \quad , \qquad (2)$$

then according to Fig. 2 we have

$$r_1 = r - \tfrac{1}{2} d \sin\theta \quad , \quad r_2 = r + \tfrac{1}{2} d \sin\theta \quad , \quad r_2 - r_1 = d \sin\theta. \qquad (3)$$

Then Eq. (1) becomes

$$\psi_c(r,\theta,t) = A(r)e^{i(\omega t - kr)} \left[e^{\frac{1}{2}ikd \sin\theta} + e^{-\frac{1}{2}ikd \sin\theta} \right] \qquad (4)$$

$$= A(r)e^{i(\omega t - kr)}\, 2 \cos\left(\tfrac{1}{2} kd \sin\theta\right) \quad . \qquad (5)$$

Taking the real part we obtain

where $$\psi(r,\theta,t) = A(r,\theta) \cos(\omega t - kr) \quad , \qquad (6)$$

$$A(r,\theta) = 2\, A(r) \cos\left[\tfrac{1}{2} kd \sin\theta\right] = 2\, A(r) \cos\left[\frac{\pi d \sin\theta}{\lambda}\right] . \qquad (7)$$

Let us check that Eq. (7) has the expected dependence on angle: At $d \sin\theta$ equal to 0, λ, 2λ, etc., we easily see that $A(r,\theta)$ takes on the values $2\,A(r)$, $-2\,A(r)$, $2\,A(r)$, etc. At $d \sin\theta = \pm 1/2\,\lambda$, $\pm 3/2\,\lambda$, etc., $A(r,\theta)$ is zero. Thus Eq. (7) has the right behavior at the interference maxima and at the nodes.

We have learned that for all kinds of traveling waves, the time-averaged intensity (energy flux) is proportional to the square of the wave amplitude averaged over one period of the motion. Thus according to Eqs. (6) and (7) we have

$$I(\theta) \equiv \langle \psi^2(r,\theta,T) \rangle = A^2(r,\theta) \langle \cos^2(\omega T - kr) \rangle = \tfrac{1}{2} A^2(r,\theta) \quad , \tag{8}$$

$$I(\theta) = 2\, A^2(r) \cos^2 \left[\frac{\pi d \sin \theta}{\lambda} \right] \quad , \text{ sources in phase,} \tag{9}$$

where the replacement of t by T indicates that we average over one period T, and where $I(\theta)$ stands for the time-averaged intensity as a function of θ with r held constant. In Fig. 3 we plot $I(\theta)$ in the region near $\theta = 0$ under the assumption that the sources are separated by many wavelengths, so that $I(\theta)$ goes through many maxima and minima while θ is still rather small. That enables us to make a diagram in which we show several maxima and minima in the same small region (near $\theta = 0$).

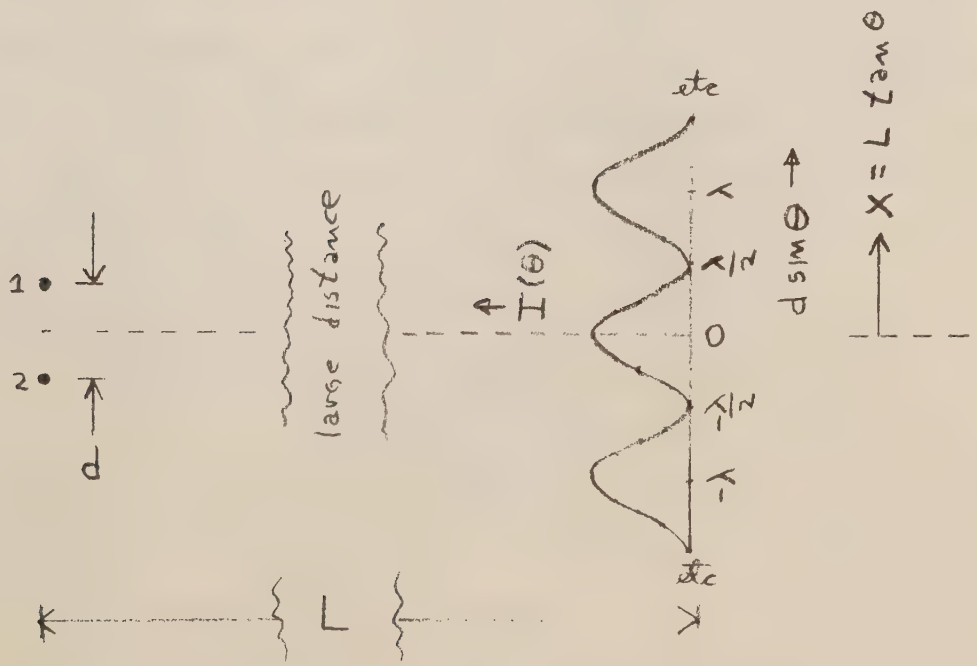


Fig. 5-3. Sources oscillating in phase. Intensity of superposition. The separation d is large compared with λ.

Total wave for sources oscillating 180° out of phase. If the sources oscillate 180° out of phase, we can replace the plus sign by a minus sign in Eqs. (1) and (4). Then in Eq. (5) the expression 2 cos is replaced by 2i sin. Taking the real part of ψ_c, we find Eqs. (6) and (7) to be replaced by the two equations

$$\psi(r,\theta,t) = -A(r,\theta) \sin(\omega t - kr) \quad , \tag{10}$$

$$A(r,\theta) = 2A(r) \sin\left[\tfrac{1}{2} kd \sin\theta\right]$$

$$= 2A(r) \sin\left[\frac{\pi d \sin\theta}{\lambda}\right] \quad . \tag{11}$$

The intensity $I(\theta)$ is given by $\frac{1}{2} A^2(r,\theta)$, which gives

$$I(\theta) = 2A^2(r) \sin^2\left[\frac{\pi d \sin\theta}{\lambda}\right] \quad , \quad \text{sources } 180^\circ \text{ out of phase.} \tag{12}$$

In Fig. 4 we plot Eq. (12) near $\theta = 0$ for the case where d is many wavelengths, so that several maxima of $I(\theta)$ occur near $\theta = 0$.

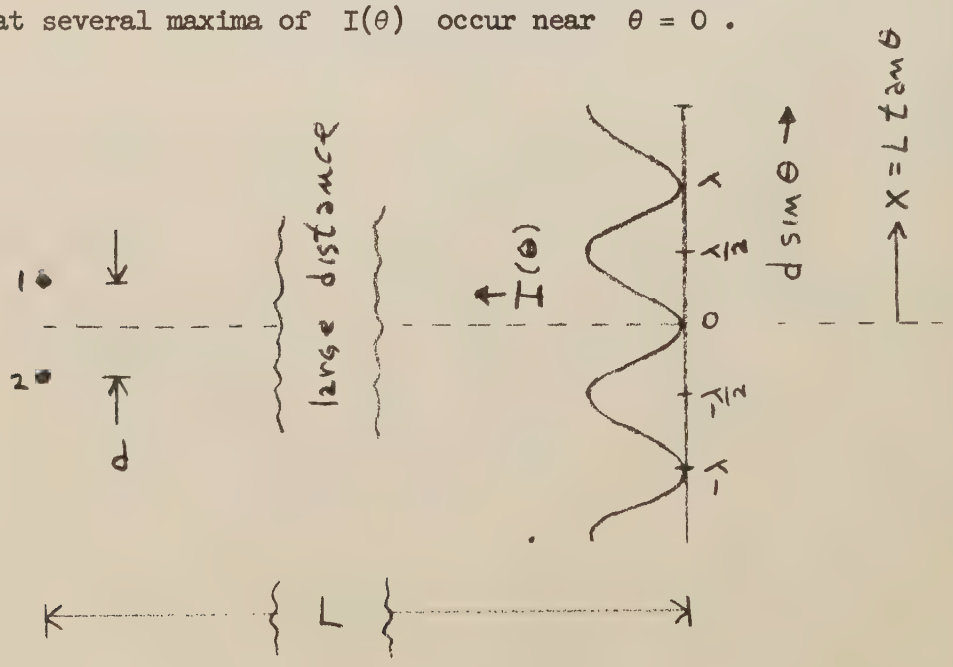


Fig. 5-4. Sources oscillating 180° out of phase. Intensity of superposition. The separation d is large compared with λ.

Energy conservation. If source 2 is turned off, the wave at P is given by

$$\text{source 1 only:}\quad \psi(r,\theta,t) = A(r)\cos(\omega t - kr_1)\ , \tag{13}$$

with corresponding intensity

$$I(\theta) = \langle\psi^2(r,\theta,T)\rangle = \tfrac{1}{2}A^2(r)\ , \tag{14}$$

which is independent of θ. When both sources are turned on, the intensity is given by Eq. (9) if the sources are in phase and by Eq. (12) if they are 180° out of phase. In either case, at interference maxima $I(\theta)$ is $2A^2(r)$ and at minima it is zero. Therefore if there are many maxima and minima between $\theta = 0$ and $\theta = 360^\circ$, $I(\theta)$ will equal $A^2(r)$ on the average. This is just twice the intensity for one source alone, as given by Eq. (14). We thus see that the total energy being emitted (in the plane of the figure) is just the sum of what the sources would give individually, provided that the two sources are many wavelengths apart. That seems reasonable.

Now consider the case where the two sources are very close together. Let them be at practically zero separation, i.e., at a separation very much less than one wavelength. If the sources are in phase, then Eq. (9) tells us that as $d/\lambda \to 0$, $I(\theta)$ becomes $2A^2(r)$, independent of angle. It is as if there were a single source which has twice the amplitude of either source alone and which therefore emits four times as much energy as does either source alone. Thus, instead of having an amount of energy that is the sum of what the two sources give individually, we get twice that amount. This may seem strange. Doesn't it violate energy conservation? No. The implication is that each source emits twice as much energy when the other source is sitting on top of it (oscillating in phase) as it does when oscillating by itself. How can that be? We have prescribed the velocities of

the sources as given, fixed quantities. It must be that the impedance experienced by each source has doubled! Why is that? It is because the "resistive drag" force exerted on the electrons in one antenna by the radiated field(taking the two radio antennas as an example) is not due just to the field being emitted by that antenna; it is that force plus the force due to the field being emitted by the other antenna. Since the currents are in phase (by hypothesis) and since the antennas are in very close proximity, the net drag force exerted on the electrons in one antenna is twice what it would be if the other antenna were not present. The power supply must therefore push twice as hard to maintain the prescribed velocity. Thus we get twice the work done by the power supply. This holds for each antenna. Hence we have accounted for the twofold increase in total energy emission.

If the sources oscillate 180° out of phase, then if you superpose one antenna almost on top of the other, you get almost zero for the total wave amplitude. In the limit that the antennas are on top of one another, the output is zero. The power supply does no work and no energy is radiated. The field emitted by one antenna pushes on the electrons in the other antenna in a way so as to help the oscillator. In the limit of zero separation of the antennas, the electrons in the two antennas drive one another with no help from the oscillator. We then have a "closed" system with energy going out of one antenna into the other and back again. The antennas are then just part of the resonant circuit of the oscillator, and the power supply need only replenish the losses due to the resistance of the antennas. The radiation "resistance" -- the characteristic impedance -- has gone to zero.

*Dipole pattern. Suppose that the sources oscillate 180° out of phase, so that the amplitude $A(r,\theta)$ is given by Eq. (11). Suppose also that d is very small compared with λ. Then $\pi d \sin\theta/\lambda$ in Eq. (11) is small compared with 1 radian for all θ. Then we may approximate the sine of this quantity by the quantity itself, neglecting higher order terms in the Taylor's series expansion of $\sin[\pi d \sin\theta/\lambda]$. Then Eq. (11) becomes

$$A(r,\theta) = A(r)\left[\frac{2\pi\ d\ \sin\theta}{\lambda}\right] = A(r)\ kd\ \sin\theta\ ,\ kd \ll 1. \qquad (15)$$

The pattern given by Eq. (15) is called a dipole pattern. It is plotted versus θ in Fig. 5 in a polar plot, i.e., a plot in which the absolute magnitude of $A(r,\theta)$ is given by the distance from the origin at azimuth θ of the point on the curve. The sign of $A(r,\theta)$ for θ between 0 and 180° is opposite to that for θ between 180° and 360°. This is indicated by the + and - signs on the sketch. Each lobe is a circle.

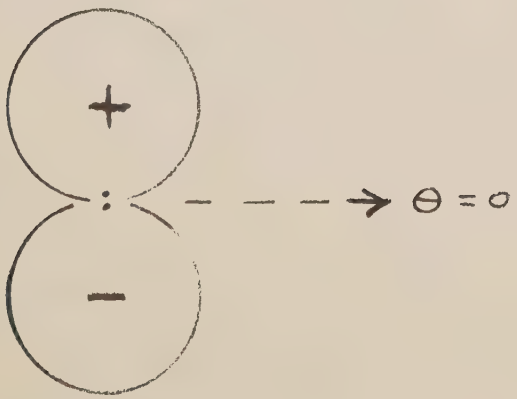


Fig. 5-5. Dipole pattern. The two sources oscillate 180° out of phase and are much closer together than one wavelength.

5.3 Interference between two independent sources.

Suppose that each of two sources has dominant angular frequency ω_o and bandwidth $\Delta\omega$. Suppose further that the sources are independent. That is, they are not driven by a common driving force. Thus there is

nothing that keeps them exactly in phase. In the case of the two radio antennas, that would mean that each antenna is driven by a separate oscillator and power supply. In the case of the sources of visible light, it means we have two independent sources with different atoms contributing to each source. For example, we may have a mercury vapor lamp consisting of a gas discharge in a glass tube which is surrounded by an opaque jacket pierced by two pin holes or slits. Each pinhole is illuminated by different atoms of the gas. Alternatively we may have two pinholes or slits in an opaque piece of material set in front of an ordinary light bulb. (In order to have a reasonably small band of frequencies, we can put a piece of red cellophane over the slits to act as a "filter.")

We shall suppose that the frequency bandwidth $\Delta\nu$ is small compared with the dominant frequency ν_o. Then there are many oscillations at frequency ν_o during a time interval of length $(\Delta\nu)^{-1}$. The time interval $(\Delta\nu)^{-1}$ is the "coherence time," t_{coh}. It is the time interval required for frequency components at extremes of the frequency band to get out of phase by about 2π. Thus if t_{coh} is defined by

$$\Delta\omega\, t_{coh} \approx 2\pi \ , \tag{16}$$

we see that t_{coh} is $2\pi/\Delta\omega$, i.e., t_{coh} is $(\Delta\nu)^{-1}$. For time intervals smaller than $(\Delta\nu)^{-1}$ we can think of the relative phase of the two sources as remaining essentially constant. (There can be many oscillations in such a time interval, because we assume $\nu_o(\Delta\nu)^{-1}$ is large.)

Let us consider only the situation in which the separation d between the two sources is large compared with the wavelength λ. Then the interference pattern looks like Fig. 3 for times when the relative phase of the

two sources happens to be zero. It looks like Fig. 4 at times when the relative phase happens to be 180°. For relative phases between 0° and 180° the interference pattern lies between those shown in Figs. (3) and (4).

If the detector is one which requires a long time to detect the intensity at a given position, as for example a photographic emulsion, then the plot of time-averaged intensity versus θ will show no θ dependence, because during a time long compared with $(\Delta\nu)^{-1}$ the interference pattern will have taken on all appearances between the extremes given by Figs. 3 and 4, and every value of $d \sin\theta$ (or of $x = L \tan\theta$) will have experienced the same time-averaged intensity. One then says that the two point sources are "incoherent." This is shown in Fig. 6.

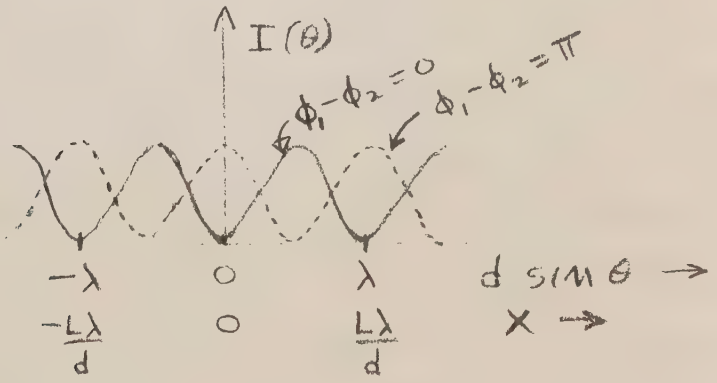


Fig. 5-6. Interference pattern from two "incoherent" sources. The relative phase $\phi_1 - \phi_2 = 0$ occurs just as frequently as the relative phase $\phi_1 - \phi_2 = \pi$ in a time interval long compared with $(\Delta\nu)^{-1}$. The average of $I(\theta)$ for these two phase differences is independent of θ. Similarly for every other phase difference ϕ, the phase difference $\phi + \pi$ occurs just as frequently, and $I(\theta)$ averages to a constant.

It is clear that there are no "intrinsically" incoherent sources. "Incoherence" is merely the result of a measurement process that throws away the information that is available in the interference pattern if one has the technique to look at times comparable to or shorter than $(\Delta\nu)^{-1}$. For visible light, the coherence times are of order 10^{-9} to 10^{-8} sec (for source consisting of independently radiating atoms in a gas discharge tube), so that it takes some experimental ingenuity to measure the interference pattern before it changes. Nevertheless it has been done in a very beautiful experiment.*

Brown and Twiss stellar interferometer. Suppose that the two sources 1 and 2 are the two members of a double star. The separation between the stars is d, and the distance from the stars to the earth is L. The angular separation between the stars as seen from earth is d/L. Suppose the two stars cannot be resolved with the largest telescope. Nonetheless we resolve them by using the trick (invented by Brown and Twiss) of looking at the interference pattern rapidly, before it changes. (Later we shall tell you how they actually "look" in a time interval of 10^{-8} sec.) Suppose that the x displacement between successive interference maxima in Fig. 3 or Fig. 4 can be measured, and the result is found to be x_o. According to Fig. 3, the x displacement is $L \tan\theta$, which we may take to be $L\theta$ since the angles we are discussing are very small. Thus a measurement of the distance x_o between interference maxima corresponds to a measurement of $L\theta_o$, where θ_o is the angular separation between successive interference maxima. Now, between two successive maxima, $d \sin\theta$ increases by one wavelength λ; i.e.,

---

* R. Hanbury Brown and R. O. Twiss, "The Question of Correlation between Photons in Coherent Light Rays," Nature 178, 1447 (1956).

for small θ, $d\theta$ increases by λ. Thus $d\theta_o$ is λ. Thus the measured distance x_o gives

$$x_o = L\theta_o = L\frac{\lambda}{d} . \qquad (17)$$

The desired quantity d/L is thus given by λ/x_o.

When Brown and Twiss invented their interferometer, they tested it by making the first measurement of the diameter of Sirius.* [Think of the uniform circular disc of a star as divided into two halves. Then think of each half as, crudely speaking, a "point." The separation of the two average "points" which crudely represent the disc of the star is about one stellar radius. Thus you can use Eq. (17), as long as you do not worry about little factors like 2 or π. Of course Brown and Twiss calculated a more exact and more complicated formula than Eq. (17).] They used light of wavelength about 4000 A^o and found a value for x_o of about 3 meters. That means they were measuring angular separations of d/L of about

$$\frac{d}{L} \approx \frac{\lambda}{x_o} = \frac{4 \times 10^{-5} \text{ cm}}{300 \text{ cm}} \approx 10^{-7} \text{ radian} ,$$

which is like measuring the size of a light bulb at a distance of 700 miles.

The method by which Brown and Twiss effectively "read" the interference pattern in a time less than 10^{-8} sec is as follows: One uses two photomultipliers at different values of x and with a variable separation $x_1 - x_2$. The output current of one photomultiplier, I_1, is multiplied by that of the other, I_2, in a fast circuit which can follow current fluctuations that occur in times of order 10^{-8} sec. (In other words the fast circuitry has 100 Mc bandwidth.) The product is determined "instantaneously," i.e., at the 10^{-8} sec level, but then the average of this product, $\langle I_1 I_2 \rangle$, is taken over

*R. H. Brown and R. O. Twiss, "A Test of a New Type of Stellar Interferometer on Sirius," Nature 178, 1046 (1956).

a long time interval of minutes. The separation $x_1 - x_2$ of the two photomultipliers is varied and the time average of the product of currents is taken at each separation. Finally one plots the time-averaged product versus $x_1 - x_2$. Now, the instantaneous current in each photomultiplier is proportional to the flux of light energy, i.e., to $I(\theta)$ at that photomultiplier First let us consider the case where the separation $x_1 - x_2$ is zero, so that each photomultiplier is subject to the same instantaneous light flux. Let us perform a very crude average of the product of the two currents. Let us say that $I(\theta)$ only takes on the four values indicated by a,b,c, and d in Fig. 7.

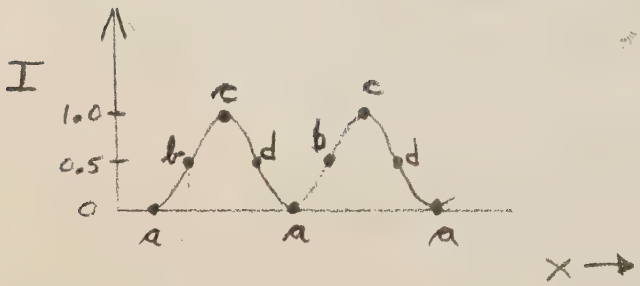


Fig. 5-7. Intensity versus x at a given "instant" of duration $\approx (\Delta\nu)^{-1}$.

Let us call the corresponding currents by the names a, b, c, and d and give them units in which we will have a = 0, b = $\frac{1}{2}$, c = 1, and d = $\frac{1}{2}$. For one fourth of the time, PM 1 (photomultiplier 1) has current I_1 corresponding approximately to a. At the same time I_2 is also equal to a, since PM 2 is at the same place as PM 1. One fourth of the time, each has current corresponding to b, one fourth of the time to c, and one fourth of the time to d as the interference pattern shifts. Thus the time average of the product of the two currents is given (in our crude approximation) by

$$x_2 - x_1 = 0 : (I_1 I_2)_{avg} = \tfrac{1}{4}(a\,a + b\,b + c\,c + d\,d)$$

$$= \tfrac{1}{4}(0\cdot 0 + \tfrac{1}{2}\cdot\tfrac{1}{2} + 1\cdot 1 + \tfrac{1}{2}\cdot\tfrac{1}{2}) = 3/8 \quad (18)$$

Now let us find the average of $I_1 I_2$ when the separation $x_2 - x_1$ is that between an interference maximum and the neighboring minimum, i.e., when it is $\frac{1}{2} x_o$, with x_o given by Eq. (17). Then when PM 1 has current a , PM 2 has b , when PM 1 has b , PM 2 has c , etc., as we see from Fig. 7. Thus

$$x_2 - x_1 = \tfrac{1}{2} x_o : (I_1 I_2)_{avg} = \tfrac{1}{4} (a\,b + b\,c + c\,d + d\,a)$$

$$= \tfrac{1}{4} (0 \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot 1 + 1 \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot 0) = \tfrac{1}{4} . \quad (19)$$

We see that $(I_1 I_2)_{avg}$ is 1.5 times larger when $x_2 - x_1$ is zero than when it is $\frac{1}{2} x_o$. Thus we see that a plot of $(I_1 I_2)_{avg}$ versus $x_2 - x_1$ will determine x_o .

The crux of the technique of Brown and Twiss is that in the product $I_1 I_2$ each current is only averaged over times of order 10^{-8} sec., and during this time the currents are essentially constant. The average $\langle I_1 I_2 \rangle$ over a time interval of minutes is just what they would get by averaging over several dozen coherence times, say over 10^{-6} sec. (They average over much longer times so as to average out photomultiplier noise and for other experimental reasons.) The product $\langle I_1 \rangle \langle I_2 \rangle$, on the other hand, is independent of $x_1 - x_2$, because each photomultiplier has sampled the entire interference pattern during the time of averaging. The essential thing is to find out which separations $x_1 - x_2$ correspond to the situation that I_1 is large when I_2 is large (as when $x_1 - x_2$ is zero) and which separations correspond to I_1 small when I_2 is large (as when $x_1 - x_2$ is $\frac{1}{2} x_o$) . In terms of photons, one asks whether the probability for detecting a photon in photomultiplier 2 is larger than average when photomultiplier 1

has "recently" (within 10^{-8} sec) detected a photon (as is the case for $x_1-x_2 = 0$) or is instead smaller than average (as for $x_1-x_2 = \frac{1}{2} x_o$) .

5.4 How large can a "point" light source be?

In Fig. 1 we showed how one can obtain two coherent sources of light (two sources whose relative phase remains constant) by irradiating two slits in an opaque screen with radiation from a "point" source of light. On the other hand, if the source is so broad that one slit is illuminated mostly by one set of atoms and the other by another independent set, then the two slits are completely incoherent, i.e., their phases are uncorrelated [for measurement times that are long compared with $(\Delta\nu)^{-1}$]. These two extremes are illustrated in Fig. 8.

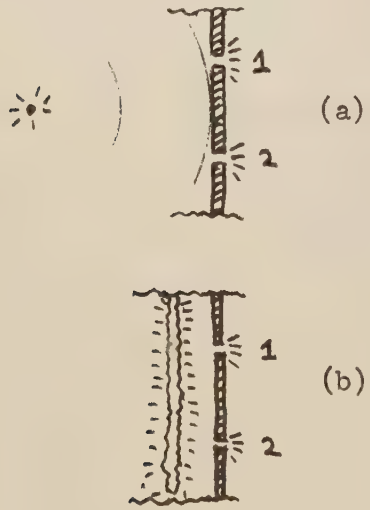


Fig. 5-8. (a) Sources 1 and 2 are driven by a single point source and maintain constant relative phase. They are coherent. (b) Sources 1 and 2 are driven by different sets of independently radiating atoms. For measurement times long compared with $(\Delta\nu)^{-1}$, they are incoherent.

The closest we can get to having a point source is to have a single atom. According to the classical picture this atom emits electromagnetic waves in all directions and drives the slit currents in Fig. 8(a) with equal phase. (The quantum theory gives effectively the same result.) A practical source of light will have a huge number of radiating atoms. If they were all sitting at exactly the same point, we would have a point source. (It would be even more like a classical point source than is a single real atom.) But in any practical

source the atoms are in a region of finite dimensions. How large can a light source be and still be an "effective" point source (meaning that the slit currents in the double slit irradiated by the "point" source maintain constant relative phase)?

Let us consider a very simple source that is not a point source. It consists of three point sources a, b, and c, each with the same dominant frequency, bandwidth, and average intensity, arrayed as shown in Fig. 9.

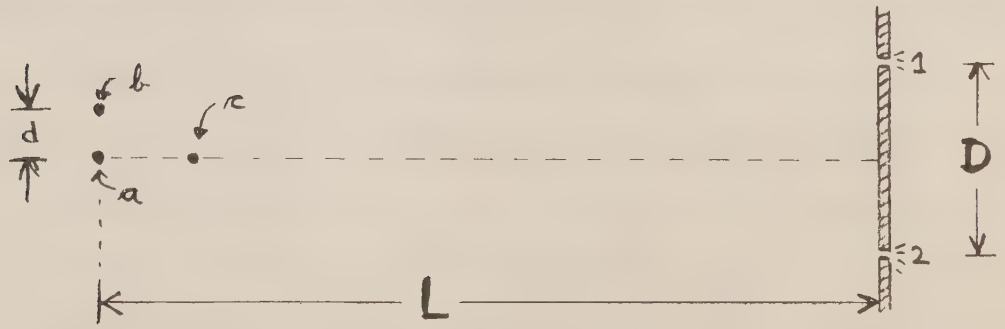


Fig. 5-9. Coherence. Slits 1 and 2 are driven by the independent sources a, b, and c. What is the condition that the relative phase of slits 1 and 2 is always essentially zero?

Suppose we start out with only point source a turned on. Then slits 1 and 2 are driven at constant relative phase (which happens to be zero for our figure) and are coherent over any time interval. Next turn on both sources a and c. Source c is a light source with the same frequency and bandwidth as source a but not correlated in phase with source a. Thus c and a do not maintain constant relative phase over times long compared with $(\Delta\nu)^{-1}$. Nevertheless the relative phase of slits 1 and 2 remains zero for all time, because source c drives the slit currents with zero relative phase just as does source a. The slit currents may be

regarded as a superposition of currents induced by the two sources, and if each source contribution gives zero relative phase between the slit currents, so does the superposition. Thus we conclude that we can extend the point source along the line connecting a and c without spoiling the coherence of slits 1 and 2 .

Now consider the situation where both sources a and b are turned on (with c turned off). Sources a and b are independent sources having the same dominant frequency and bandwidth and the same average intensity. During any time interval short compared with $(\Delta\nu)^{-1}$, the amplitude and phase constant of each source remain constant. Suppose that a given instant (an instant means a time interval short compared with the coherence time $(\Delta\nu)^{-1}$ but long enough to have many fast oscillations, so that we can tell what the amplitudes and phases are) it so happens that the amplitude of b is very small compared with that of a . Then to a good approximation the two slits are irradiated only by a , and the slit currents therefore have zero relative phase. Now let us wait a time long compared with the coherence time of sources a and b and look again. Suppose that this time it so happens that the amplitudes of the oscillations of a and b are practically equal. In that case the screen with the slits is irradiated by the two-source interference pattern that we have seen in Figs. 3, 4, 6, and 7. The locations of the maxima and minima depend on the relative phase of sources a and b . The question of interest is whether or not the two slits 1 and 2 are still driven with zero relative phase. We know that the amplitude of the interference pattern changes sign when we go from one interference maximum to the next. (According to Eqs. (7) and (11), the amplitude $A(r,\theta)$ is proportional to the cosine or the sine of $\pi d \sin\theta/\lambda$,

depending on whether the sources are in phase or 180° out of phase. In either of these cases or for any other relative phase of the sources, the path difference $d \sin \theta$ increases by an amount λ between successive interference maxima, and $A(r,\theta)$reverses sign.) We now see that both slits are driven at zero relative phase most of the time only if they separated by less than the separation between successive interference maxima of the two-source interference pattern. (Even when the slits are closely spaced, it may happen that a zero of the two-source pattern irradiating them falls between the two slits, in which case they are driven 180° out of phase. However, this happens a smaller and smaller fraction of the time as the slits get closer together.)

The quantitative expression of the above result is as follows: (For simplicity we use small-angle approximations. Thus $d \sin \theta$ becomes $d \cdot \theta$, with θ in radians.) Between successive maxima, $d \cdot \theta$ increases by λ. We want the angle θ_{12} between the "ray" from the source (either source) to slit 1 and the "ray" to slit 2 to be less than the increment of θ between successive interference maxima. Thus we want

$$d \cdot \theta_{12} < \lambda \quad . \tag{20}$$

But the angle θ_{12} is given (using small-angle approximations) by the separation D between the two slits divided by the distance L from the source to the slits:

$$\theta_{12} = \frac{D}{L} \tag{21}$$

Combining Eqs. (20) and (21), we obtain the condition for coherence: dD/L must be less than the wavelength. An interesting way to express this result

is to notice that the product dD of the source separation and slit separation is less than the product of the "large" distance L and the "tiny" distance λ :

coherence condition: $\boxed{d\,D < L\,\lambda}$. (22)

If the source consists of a huge number of points between a and b , so that the source has width d , Eq. (22) applies to the entire source if it applies to the extreme points a and b (i.e., point sources closer together than d are coherent if those d apart are). Similarly, when we (later) consider several or many slits in a screen instead of just two, the coherence condition Eq. (22) can be applied to the entire array of slits with D taken as the separation of the outermost slits.

You can verify Eq. (22) in an easy experiment. (See Home Experiment 17.)

5.5 Angular width of a "beam" of traveling waves -- approximate treatment.

A "beam" of traveling waves is a pattern of waves traveling in a given direction and having a finite lateral width. A flashlight beam of visible light and a radar beam of microwaves are each made by putting a small source of electromagnetic radiation at the focal point of a parabolic reflector. The small source drives the electrons in the metallic surface of the reflector with just the proper phase relations so that the reflected radiation from all points of the surface interferes constructively along the direction of the beam. Another way to get a light beam is to reflect light from a small or distant source (the sun, for example) off a small plane mirror. Alternatively, we can use a hole in an opaque screen instead of a mirror. If the source is sufficiently far away and sufficiently small,the radiation incident on the mirror (or hole) can be approximated as a plane wave -- i.e.,

one with all the radiation traveling in exactly the same direction. Then the mirror reflects "part of the plane wave." Similarly in the case of the small source at the focus of a parabolic mirror, if the source is sufficiently small and the mirror is a perfect paraboloid, the beam is (to a certain approximation) like a "segment of a plane wave," consisting of radiation all traveling in the same direction. All of these considerations hold as well for sound waves or water waves.

Now comes the interesting question: Can one, by very careful design, make a beam of waves that is just like a "cross sectional segment" of a plane wave, in the sense that all the waves are traveling in exactly the same direction, so that one has a perfectly parallel beam that will continue forever with the same width? No. No matter how small the point source at the focus of a perfect parabola, the radiation in the beam will not be perfectly parallel. If the "dominant" direction is along $\hat{z}$ and the width of the beam (at a given value of z, say at the reflector) is D, then the angular distribution of propagation directions will have a "full width at half maximum intensity" of about λ/D. (We will show this below.) Similarly if we have a plane wave from a distant point source falling on a hole of width D (or a mirror of width D), the angular width of the beam is about λ/D. The angular width can only be zero if D is infinite (or if λ is zero). The angular width of the beam is said to be "diffraction limited." In Fig. 10 we show some examples of beams. Note that if the original width of the beam is D, and if every attempt is made to make the beam as perfectly parallel as possible, the width W after the beam has traveled a large distance L is approximately the original width D plus L times the angular fill width λ/D. For large enough L, we can neglect the original width D. Thus we have

angular full width: $\boxed{\Delta\theta \approx \frac{\lambda}{D}}$ (23)

beam width: $W \approx L\frac{\lambda}{D}$ (24)

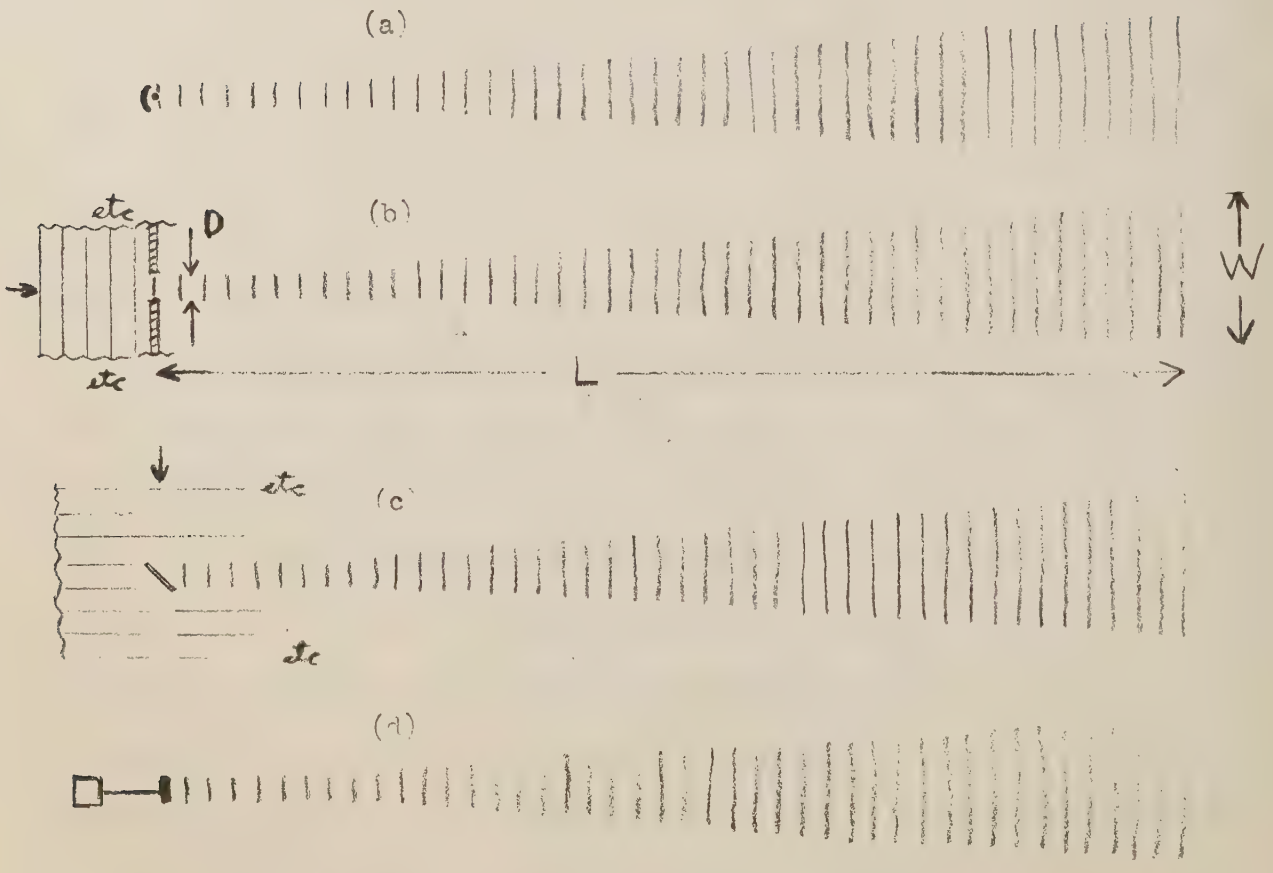


Fig. 5-10. Diffraction. Beam of width D has angular width $\approx \lambda/D$ and spreads by an amount $W \approx L\,\lambda/D$ in traveling a distance L. (a) Beam made by point source and parabolic mirror. (b) Beam made by plane wave incident on hole in opaque screen. (c) Beam made by plane wave incident on plane mirror. (d) Beam emitted by plane radiator.

Every one of the four sketches in Fig. 10 can be taken to represent either water waves, sound waves, or electromagnetic waves (visible light, of

wavelength 5×10^{-5} cm, for example, or microwaves of wavelength 10 cm, for another example).

Here is a crude derivation of Eq. (23). (In Sec. 5.6 we shall give an exact derivation.) The result is independent of the kind of waves and is independent of how the waves are produced. We might as well take the simplest source, which is probably a plane radiator, as shown in Fig. 10(d). For sound waves this can be an oscillating piston in free air. For electromagnetic waves it can be an oscillating sheet of charge of finite extent, for example a plane antenna array. In any case the entire radiator is coherent. That is, all the "moving parts" move in phase with one another. In the dominant direction of the beam, a field point far from the radiator is equidistant from all parts of the radiator. Therefore waves from all parts of the radiator add with the same relative phase and we have a constructive interference maximum. That is what defines the dominant direction of the beam. (If one varies the relative phase over the surface of the radiator one can "steer" the beam in a direction that is not normal to the surface of the radiator. That is exactly what happens in Fig. 10(c), where different parts of the mirror tilted at 45° to the incident plane wave are driven with different phase, so that the region of maximum constructive interference -- the direction of the reflected beam -- is not normal to the mirror but instead satisfies the law of "specular reflection.")

At a distant field point that is not quite in the direction of the beam one does not have completely constructive interference. To see where we get the first zero in the interference pattern, let us divide the radiator into two halves, a top half and a bottom half. Then we approximate the radiator by two coherent point (or line) sources, one at the middle of the

top half and one at the middle of the bottom half. These sources have lateral separation $\frac{1}{2}$ D . Their first interference zero (the first zero on either side of the principal maximum along the beam direction) occurs for a path-length difference of half a wavelength, i.e., when $(\frac{1}{2}$ D) sin θ is $\frac{1}{2}\lambda$. For small angles we take sin $\theta = \theta$ and have

$$\text{half angular width to first zero} = \frac{\lambda}{D} \qquad (25)$$

This is shown in Fig. 11.

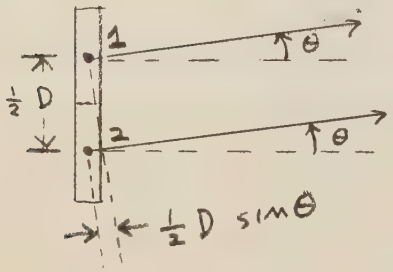


Fig. 5-11. Plane radiator. Source 1 represents the contributions from the top half, source 2 those from the bottom half.

The result Eq. (25) is exact. Where does the next maximum occur? If points 1 and 2 of Fig. 11 really were point (or line) sources, the next maximum would occur when the path length from the field point to source 2 exceeds that to source 1 by one wavelength. Indeed the top half and bottom half are in phase then, but they each contribute zero! That is because if you divide the top and bottom halves themselves into halves, so that the entire radiator is divided into four quarters, then the contribution of the first quarter is 180° out of phase with that of the second and cancels it; that of the third is 180° out of phase with that of the fourth and cancels it. Thus the first subsidiary maximum actually comes not when we have two halves with contributions differing in phase by 2π , i.e., four quarters with contributions differing in phase between successive quarters

by π , but instead when we have three thirds of the radiator differing in phase by π from one third of the radiator to the next. Two of the thirds cancel one another, but the third third remains. Thus the amplitude for the first subsidiary maximum is smaller than that of the principal maximum by at least a factor of 1/3 (it is actually smaller by more than that because of phase differences within the one-third that is left). We see that the subsidiary maxima have small amplitude compared with the central maximum that gives the "beam" direction. When we study the exact pattern we will find that the half angular width to the first zero is equal to the full angular width at about half maximum intensity, which is how we have defined the angular width of the beam in Eq. (23). Thus we have derived Eq. (23), roughly.

The result Eq. (23) can be put in a different interesting form. Let us think of a single frequency component of the traveling wave. We may take this component to be exactly monochromatic. Then the bandwidth $\Delta\omega$ is zero. What about the propagation vector? The square of the propagation vector, k^2 , is equal to ω^2/c^2 (for light in vacuum). Therefore k^2 must have a perfectly definite value if ω^2 does. But that does not mean that the components of $\underset{\sim}{k}$ must have definite values. Now, k^2 is the sum of the squares of its components

$$k^2 = k_x^2 + k_y^2 + k_z^2 \quad , \tag{26}$$

where k_x gives the number of radians of phase per unit distance along $\hat{x}$, k_y gives radians of phase per unit distance along $\hat{y}$, and k_z gives radians of phase per unit distance along $\hat{z}$. If the beam were a plane wave traveling along $+z$, then k_x and k_y would be zero. For a Fourier component of the beam that is traveling with a propagation vector in the x-z

plane and making a small angle θ with the z axis, we have $k_y = 0$, $k_x = k \sin\theta$, and $k_z = k \cos\theta$. For small angles θ, we can approximate $\sin\theta$ by θ and $\cos\theta$ by 1. Then we have for the x component

$$k_x \approx k\,\theta \quad . \tag{27}$$

But we have already seen that the "beam" has an angular spread about the dominant direction $\hat{z}$ given by [see Eq. (23)]

$$\Delta\theta \approx \frac{\lambda}{D} \quad . \tag{28}$$

Thus the spread in k_x is given by [combining Eqs. (27) and (28)]

$$\Delta k_x \approx k\,\Delta\theta \approx k\,\frac{\lambda}{D} = \frac{2\pi}{D} \quad ,$$

or, writing Δx for the full width D of the beam in the x direction we have

$$\boxed{\Delta k_x\,\Delta x \approx 2\pi} \quad . \tag{29}$$

For comparison we recall the results of Sec. 2.4, where we found

$$\Delta k_z\,\Delta z \approx 2\pi \quad , \tag{30}$$

$$\Delta\omega\,\Delta t \approx 2\pi \quad . \tag{31}$$

Equations (29), (30), and (31), together with an analogous equation for the y direction, $[\Delta k_y\,\Delta y \approx 2\pi]$ are the "guts" of Fourier analysis.

5.6 Multiple-slit interference pattern and single-slit diffraction pattern.

In this section we shall consider the interference pattern resulting from an arbitrary number of uniformly spaced parallel slits in an opaque screen which is irradiated by a point source of light very far from the screen. All of the slits are coherent, i.e., they maintain constant relative phase. We shall take the line from the slits to the distant point source to be normal to the screen, so that all the slits are driven at the same phase. The light incident on the screen will be taken to be a plane wave. (Its polarization does not concern us. All of our considerations apply to each polarization component.)

Opaque screen. First we should understand how an opaque screen works. All electromagnetic radiation has its ultimate origin in oscillating charged particles. The total electric (and magnetic) field at any given point is a superposition of the waves produced by all of the sources, i.e., all the oscillating charges. In the present problem one of the sources is the distant point source that produces the plane wave incident on the screen. We shall call this the source S . Behind the opaque screen the total wave amplitude is zero (by hypothesis -- that is what we mean by an opaque screen). This total wave is a superposition of the wave from S and the waves emitted by the oscillating electrons in the material of the screen. That is, the screen doesn't just "gobble up" the incident wave from S . Its electrons are driven by the incident radiation (and also by the radiation emitted by the other electrons of the screen), and the superposition of all of the waves, i.e., from S and from all the electrons, gives zero behind the screen. If this seems strange to you, recall how it comes about that a static electric field is zero inside of a good metallic conductor. The conductor doesn't "eat up" the external "driving" field. That field is

"still there" inside the conductor. But charges move in the conductor (before static equilibrium is established) and come to rest on the surfaces until finally the superposition of the fields from the surface charges and the incident field gives zero total field inside the conductor. All electromagnetic fields come from charged particles, and such a "zero" field is the result of superposition.

Here is a homely analogy which is often met: If you have a mental picture of the electric lines of force of a charged particle as being like a little stream of bullets sent out at the velocity of light from the point charge, you will get into serious trouble. Little bullets do not obey the superposition principle. They do not pass through one another without disturbance. Two bullets cannot superpose to give zero bullets. With this misleading mental picture, you will probably think of the effect of a metallic conductor on an electrostatic field as one of "stopping the bullets," like a sort of armor. That is also the way in which you might think of a screen which is opaque to incident light, as a sort of an armor that stops the light and eats it up, turning it into heat (if the screen is black) or bouncing the bullets back (if the screen is a shiny metal foil). This is a bad picture. If you have it, you will not be the first - but get rid of it. It is wrong.

There are two extremes in the varieties of opaque screen. At one extreme we have a shiny metallic foil (such as an opaque piece of aluminum foil). The electrons in the metal are driven by the incident radiation; consequently they emit electromagnetic waves. In the forward direction (the direction of the incident radiation), the superposition of the incident wave and that from the electrons gives zero. In the backward direction it gives a reflected wave.

The motion of a given electron is 90° out of phase with the total electric field at its location, and no work is done on the electron during any complete cycle. (The electron "redirects" the radiation energy without permanently absorbing any energy.) At the other extreme we have a black opaque screen (such as black cardboard or a microscope slide painted with a layer of "aquadag" - soot suspended in water). Again the electrons are drived by the incident radiation. They also suffer a resistive drag from the medium and are always at terminal velocity. Their radiation in the forward direction is 180° out of phase with the incident radiation and superposes with it to give zero (after sufficient thickness of screen). The velocity of a given electron is always in phase with the total electric force at its location (there is a minus sign between the force and the field because of the negative charge of the electron), and consequently some net work is done on the electron. The work done on the electron is transferred to the medium, which gets hotter. There is no net reflected wave -- the contributions from different layers of the screen superpose to give zero in the backward direction.

Now let us proceed to cut a small hole (or slit) in our opaque screen. First let us mark the material to be removed. This will be called slit number 1, so we label the material to be removed with a number 1. The material of the screen above and below number 1 is labeled a (above) and b (below). The total field behind the screen, which is zero, is the superposition of the fields emitted by the source S and by the material from a, b, and 1. Thus <u>before removing the material of slit 1</u> we have

$$E = 0 = E_S + E_a + E_b + E_1 \qquad (32)$$

This situation is shown in Fig. 12.

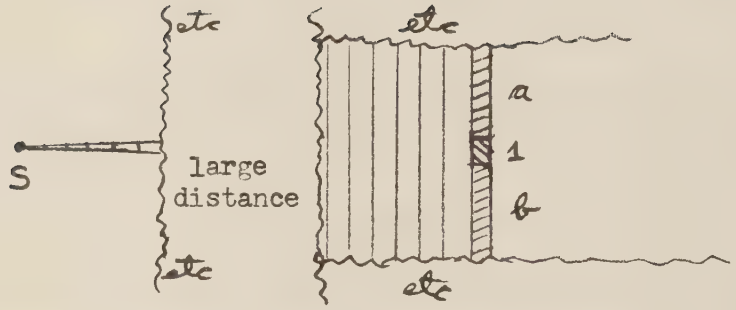


Fig. 5-12. Plane waves from distant point source S are incident on opaque screen. Superposition of fields due to charges at S, a, b, and 1 give zero behind the screen.

Now we remove the material that is plugging up slit 1. Assume that the motion of the electrons in the regions a and b is not changed by removal of the plug. (This is an approximation, because the electrons in regions a and b are driven by the total field at their locations, and that includes the fields radiated by the electrons that were in the plug. Those electrons in a and b within a few wavelengths of the edge of the slit will be most affected by removing the plug, because the radiation from a given electron falls off as r^{-1} with distance from the electron, so that the nearest neighbors are the most important.) Then the total field behind the screen is no longer the superposition given by Eq.(32). Instead it is merely that superposition minus the contribution from plug 1:

$$E = E_S + E_a + E_b = -E_1, \tag{33}$$

where the second equality in Eq.(33) made use of Eq.(32). We see that the remaining field, which is a superposition of contributions from the source S and the remaining material of the screen, a and b, is just the same (except for a minus sign) as that which <u>was</u> being emitted by the plug. Thus we can

find the field behind the screen by imagining that we substitute for the source and screen with slit the simpler system consisting of just the plug by itself, with no source S and no remaining screen, and with the electrons in the plug oscillating in the prescribed manner. This gives us an easy way to calculate the interference patterns due to slits in an opaque screen.

This calculational device is called Huygens' construction. It can be used with any number of slits and also with a single broad slit. Notice that its basis is in Eqs.(32) and(33). Equation (32) holds only behind the screen. Therefore the mentally substituted "radiating plug" gives the interference pattern only behind the screen. A real "radiating plug," i.e., an antenna, radiates in all directions. The Huygens plug only gives the right answer when used to calculate the field to the right of the screen (taking the incident radiation to be incident from the left as in the figure). Another thing to notice is that in writing Eq.(33) we assumed E_a and E_b were the same with and without the plug in place. This is only approximately true, as mentioned above. If one has, for example, a single wide slit and uses the Huygens construction to calculate the fields to the right of the screen and in the slit itself, one finds the following: If one is sufficiently far to the right of the screen and sufficiently near the forward direction, and if the screen is many wavelengths wide, then the Huygens construction gives a very good approximation to the right answer (as determined by experiment). If one is in the vicinity of the slit itself, the Huygens construction gives a very poor approximation to the right answer. If you are in the slit, the most important moving charges in the remaining screen material are those nearest the edge of the slit, because they are closest. But these are just the ones most affected by the removal of the plug. The field pattern can be very complicated in the

slit and especially near the edges of the slit, where the nearest oscillating charges dominate. You may ask, "Why not just solve the problem exactly?" That is very difficult. You must use Maxwell's equations in all of the vacuum regions and in the material, specifying the properties of the material precisely, and fitting everything together at the boundaries. There are no general methods of finding solutions, and very few problems have been solved exactly.

Interference pattern of N slits using Huygens' construction. We shall calculate the interference pattern to the right of an opaque screen having N identical and uniformly spaced slits, where N is any number from 2 to infinity. The total width of the region in which there are slits will be held constant at width D as we vary N. This width is the lateral separation of the centerlines of the first and last (Nth) slits. The separation between centerlines of adjacent slits will be called d. Thus D is (N-1)d. We shall think of each slit as being very narrow, so that we can neglect any phase shift between contributions from different parts of the same slit. In the limit that N goes to "infinity" (with D held constant), we shall have the pattern of a single wide slit of width D. (If we prefer, we can think of the slit width as being a very small constant width, a. Then we will have one wide slit when N is D/a and d is a. These details never matter, as we never really mean it literally when we say "infinity.") After we understand the interference pattern, we can, if we wish, consider the interference pattern of N slits of non-negligible width a, separated by distance d, and covering a total region of width D.

By the Huygens construction we can think of each slit as being a little "antenna" (at least for radiation to the right--though not to the left, i.e., not back towards the source). Our results will also hold for the interference

pattern of real antennas. Suppose the incident plane wave is in the +z direction. The N slits are along $\hat{x}$. By Huygens' construction we discard the source and screen and substitute N antennas along $\hat{x}$, as shown in Fig. 13.

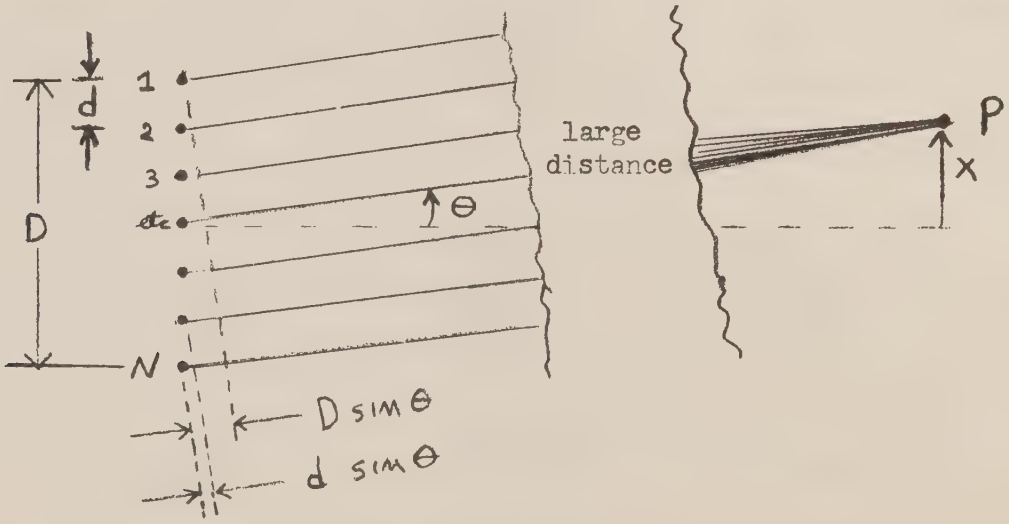


Fig. 5-13. N narrow slits (or N antennas).

At a distant field point P, each antenna gives a contribution that has the same amplitude A (because P is distant enough that the dependence of amplitude on distance is approximately the same for all the antennas). Let us use the index i to designate the ith antenna, with i = 1,2, ..., N. Then the contribution to the electric field at P from the ith antenna is the real part of

$$E_i = A\, e^{-i\omega t}\, e^{ikr_i} \quad . \tag{34}$$

(Notice that we are taking all the antenna currents to be in phase.) Let us designate the quantity kr_i by the name ϕ_i. Then ϕ_i is the phase retardation corresponding to the time delay while the wave from the ith antenna travels the distance r_i from the antenna to the field point P. The increment in this phase retardation between the contribution from antenna i and that from antenna i+1, called α, is the same for all i (since the slits are uniformly spaced) and is given by

$$\alpha \equiv \phi_{i+1} - \phi_i = \phi_2 - \phi_1 = k\,d\,\sin\theta = 2\pi\,\frac{d\,\sin\theta}{\lambda}\,. \tag{35}$$

Similarly, the difference in phase between the contribution from the Nth antenna and that from the first is

$$\phi_N - \phi_1 = k\,D\,\sin\theta = (N-1)\alpha\,. \tag{36}$$

The average phase of the contributions from the entire array is just the phase of the central antenna, if N is odd. More generally, it is the average of the phases of the first and last antennas. Calling the average of the distances r_i by the name r_{av}, we have

$$\phi_{av} \equiv kr_{av} = \tfrac{1}{2}(\phi_1 + \phi_N) = \phi_1 + \tfrac{1}{2}(\phi_N - \phi_1) = \phi_1 + \tfrac{1}{2}(N-1)\alpha\,. \tag{37}$$

The total electric field at P is the real part of the superposition

$$E_c = Ae^{-i\omega t}e^{i\phi_1} + Ae^{-i\omega t}e^{i\phi_2} + \ldots + Ae^{-i\omega t}e^{i\phi_N} \tag{38}$$

$$= Ae^{-i\omega t}e^{i\phi_1}\,[1 + e^{i\alpha} + e^{2i\alpha} + \ldots + e^{i(N-1)\alpha}] \tag{39}$$

$$= Ae^{-i\omega t}e^{i\phi_1}\,[1 + a + a^2 + \ldots + a^{N-1}]\,, \quad a \equiv e^{i\alpha}\,, \tag{40}$$

$$= Ae^{-i\omega t}e^{i\phi_1}\left[\frac{a^N - 1}{a - 1}\right] \tag{41}$$

$$= A\,e^{-i\omega t}e^{i\phi_1}\,\frac{a^{\frac{1}{2}N}}{a^{\frac{1}{2}}}\,\frac{a^{\frac{1}{2}N} - a^{-\frac{1}{2}N}}{a^{\frac{1}{2}} - a^{-\frac{1}{2}}} \tag{42}$$

$$= A\,e^{-i\omega t}e^{i\phi_{av}}\,\frac{\sin\frac{1}{2}N\alpha}{\sin\frac{1}{2}\alpha}\,. \tag{43}$$

[See the Appendix, Eqs.(21) through (34), for details omitted here. See problem for a graphical solution using the vector-diagram representation of complex numbers.] Taking the real parts, we obtain

$$E(r,\theta,t) = A(\theta)\,\cos(\omega t - kr_{av})\,, \tag{44}$$

where

$$A(\theta) = A\,\frac{\sin\frac{1}{2}N\alpha}{\sin\frac{1}{2}\alpha}\,. \tag{45}$$

We can write Eq.(45) either in such a way as to emphasize d or in such a way as

to emphasize D. Thus, using Eqs. (35) and (36), we have

$$A(\theta) = A \frac{\sin\left[N\pi \frac{d \sin\theta}{\lambda}\right]}{\sin\left[\pi \frac{d \sin\theta}{\lambda}\right]} \tag{46}$$

or

$$A(\theta) = A \frac{\sin\left[\frac{N\pi}{N-1} \frac{D \sin\theta}{\lambda}\right]}{\sin\left[\frac{\pi}{N-1} \frac{D \sin\theta}{\lambda}\right]} \tag{47}$$

Notice that for N = 2, Eqs.(44) and (46) [or Eq.(47)] reduce to the two-source result obtained in Sec. 5.2 [Eqs.(6) and (7)]. We shall let you plot $A(\theta)$ versus $\sin\theta$ for values of N such as 3, 4, etc. (Problem).

Central maximum and other principal maxima. The central maximum is the one along the direction of the incident wave, i.e., along $\theta = 0$. According to Eq.(46), the amplitude for the central maximum is NA. [We have used the fact that the limit as x goes to zero of $(\sin Nx)/x$ is N.] That is what we expect. All N of the antennas give contributions which are in phase at $\theta = 0$. Both numerator and denominator of Eq.(46) vanish at $\theta = 0$ and also when $d \sin\theta$ is λ, 2λ, 3λ, etc. It is easy to show that $A(\theta)$ is $\pm NA$ at these places. These are called principal maxima. The first principal maximum on either side of $\theta=0$ is called a first order maximum(the one at $\theta=0$ is called the zeroth order maximum). The principal maxima occur when all of the antennas give contributions that are in phase with one another, i.e., when the path length to P increases by λ from one antenna to the next. Notice that when the spacing d between antennas is less than the wavelength λ, there is only one principal maximum, the central maximum. That is true for any value of N.

Single-slit pattern. If we hold D fixed and let N become huge, i.e., let d go to zero, we obtain the diffraction pattern for a single broad slit of width D. There is only one principal maximum, the central maximum in the direction of the

incident wave. The amplitude $A(\theta)$ may be put in a reasonable form as follows: Start with Eq.(46). For N huge, N and N-1 are indistinguishable. Thus Nd may be replaced by D in the numerator of Eq.(46). In the denominator, d/λ is tiny, so that $\sin[\pi(d \sin\theta/\lambda)]$ is just $\pi(d \sin\theta/\lambda)$. Then replace d by D/N. Now recall or notice that at $\theta = 0$, $A(\theta)$ is NA. Finally we can eliminate N (which is "infinite"), d (which is "zero"), and A (which is "zero") by writing Eq.(46) in the form

$$A(\theta) = A(0) \frac{\sin\left[\frac{\pi D \sin\theta}{\lambda}\right]}{\frac{\pi D \sin\theta}{\lambda}} . \qquad (48)$$

The intensity (time-averaged energy flux) for the single-slit pattern is given by

$$I(\theta) = I(0) \frac{\sin^2\left[\frac{\pi D \sin\theta}{\lambda}\right]}{\left(\frac{\pi D \sin\theta}{\lambda}\right)^2} \qquad (49)$$

The amplitude and intensity patterns of Eqs.(48) and (49) are plotted in Fig.14.

Angular width of a beam. We have now justified the result given in Sec. 5.5 that a "beam" of width D has an angular width $\Delta\theta$ approximately equal to λ/D. The precise shapes of the amplitude and intensity versus θ are given by Fig. 14. The main feature of the intensity plot is that the intensity is large only in an angular band between $\theta = -\frac{1}{2}\lambda/D$ and $\theta = +\frac{1}{2}\lambda/D$.

Diffraction grating. Instead of N antennas or N slits in an opaque screen, we may have N parallel scratches on a piece of smooth glass or plastic of width D. If there were no scratches, the light would give a diffraction pattern corresponding to a single wide slit of width D. The scratches act like "antennas." They given an "N scratch" interference pattern which is that we just obtained for N slits, with one exception. At the central maximum (at 0°), we get con-

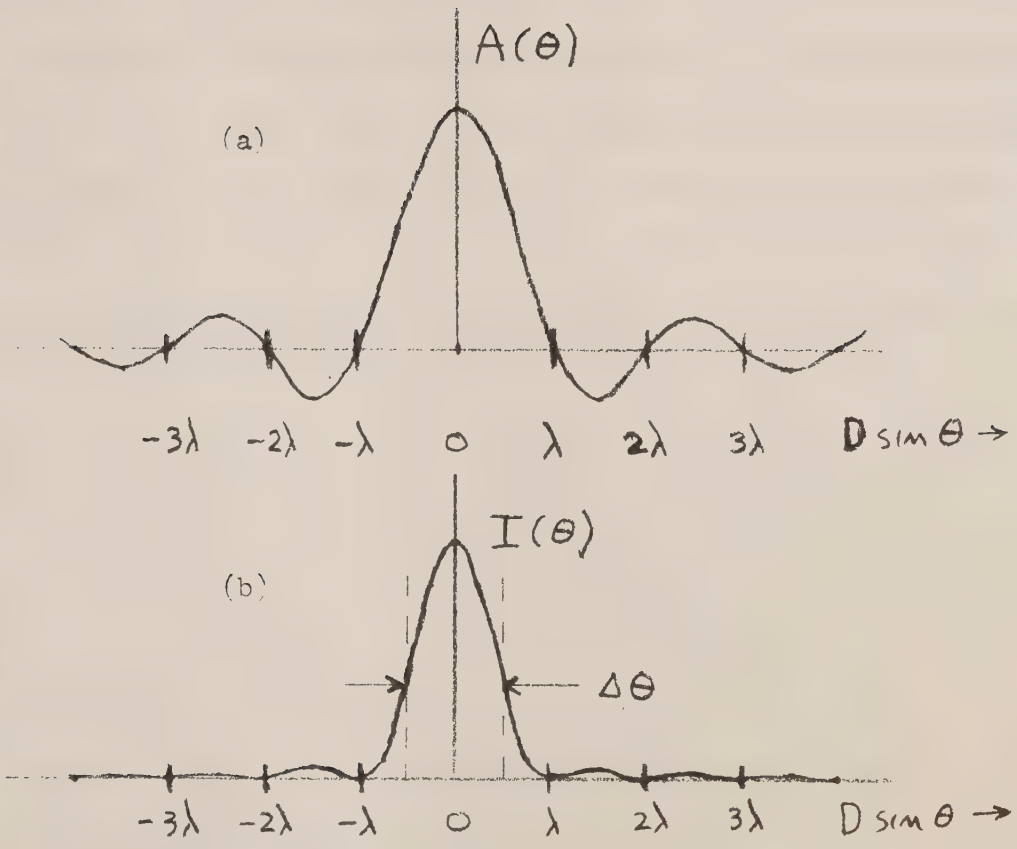


Fig. 5-14. Single slit diffraction pattern. (a) Amplitude. (b) Intensity. The angular band $\Delta\theta$ extending from $-\frac{1}{2}\,\lambda/D$ to $+\frac{1}{2}\,\lambda/D$ corresponds approximately (for small angles) to the "full width at half intensity". [The intensity is down by a factor $(2/\pi)^2 = 0.41$, rather than 0.5.]

tributions not only form the scratches but also from all the transparent material between the scratches. Therefore we expect the central maximum to be brighter than the other principal maxima by (very roughly) the ratio d/a, where d is the scratch centerline separation and a is the width of a scratch.

Monochromatic light seen through a diffraction grating has at each principal maximum an intensity profile (plot of intensity versus angle) like that of the single slit pattern of Fig. 14.

The diffraction grating in your optics kit has the above design. It has 13,400 lines per inch, which gives $d = 1.90 \times 10^{-4}$ cm., i.e., 1.9 microns. For green light with wavelength about 5500 A^{o}, i.e., 0.55 microns, how many principal maxima do you expect to find? According to Eq.(48), the principal maxima occur at values of $\sin\theta$ of 0, λ/d, $2\lambda/d$, etc. Of course $\sin\theta$ cannot be greater than 1. For our grating we have $d \approx 3.5\lambda$ when $\lambda = 0.55\mu$. Therefore if $\sin\theta = m\lambda/d$, we can have $m = 0, \pm1, \pm2$, and ±3, but not ±4. Now look at a light bulb with your grating. The "straight ahead" bulb is the central maximum. All colors overlap at $\theta = 0$. The streaks of colors at the side are light bulb images of various colors at angles corresponding to $d \sin\theta = \lambda$ for first order, 2λ for second order, etc. Do you see all three orders? (If you see four, somebody is wrong.) If you want to see the colors of an incandescent bulb as they really are, you should not use a large light bulb, because its size causes overlap of the various "colored light bulbs." Either put a narrow vertical slit over your light bulb (and hold the grating so it spreads the colors out horizontally) or, better, get a "display bulb" at any hardware or grocery store. (They have a clear glass envelope and a straight filament about three inches long, and they cost about 40 cents.)

Angular resolution of the human eye. Take a millimeter scale (or make marks on a piece of paper) and find the distance for which the lines become blurred together and are thus unresolved. Typically you will find 1 mm can be barely resolved at 2 meters and not at all at 4 meters. Thus for the human eye at the center of the field of view (i.e., looking directly at the lines) we find an

angular resolution limit $\Delta\theta \approx 1$ mm/2m = 1/2000. Now look in the mirror and measure the diameter D of your pupil, using a ruler held near the eye. Typically D ≈ 2 mm. The diffraction limit of angular resolution for your eye is given by the angular size of the image spot on your retina for an incident plane wave emitted by a distant point source. The angular full width $\Delta\theta$ of the image of a distant point is thus

$$\Delta\theta(\text{diffraction limit}) \approx \frac{\lambda}{D} \approx \frac{5.5 \times 10^{-5}\text{cm}}{0.2\ \text{cm}} \approx \frac{1}{4000} \ .$$

Thus the brain likes to have the lines separated by an angular separation of about twice the diffraction width before it sees the lines as resolved.

The actual lateral width on your retina of the image of a distant point is given approximately by the focal length of the eye's lens times the angular width of the image. The focal length f is the inner diameter of the eye, about 3 cm (when a distant object is viewed). Thus the lateral width of the image spot for a distant point is roughly $f(\lambda/D) = 3 \times 5 \times 10^{-5}/0.2 = 8$ microns. The fact that you do about as well as the diffraction limit implies that the photoreceptors at the center of the retina (these are the ones called cones) are separated by no more than about 8 microns.

An astronaut orbiting at an elevation of about 150 miles once said that he could see the individual houses in the villages as they passed beneath him. Do you believe him?

Diffraction by an opaque obstacle. In Fig. 12 we showed a point source S, and an opaque screen consisting of parts a, b, and 1. We removed the plug 1 and got the diffraction pattern from a plane wave incident on a hole, as shown in Fig. 10(b). What happens if we leave the central plug 1 and remove all the rest of the screen? Before removing anything, we still have E equal to zero and given

by the superposition [from Eq.(32)] $E_S + E_a + E_b + E_1 = 0$. Now remove a and b and assume that the motions of the electrons in plug 1 are unchanged. Then the field behind the plug is $E_S + E_1$. In the region close to the plug we know this adds up to zero; it gives the "shadow" of the plug. But of course the plug does not emit a perfect plane wave but rather a "beam" of angular full width $\Delta\theta \approx \lambda/D$ where D is the width (or diameter) of the plug. We can call this a beam of "negative light," because just behind the plug it is equal to but 180° out of phase with the incident plane wave, and thus gives a shadow when superposed with the incident plane wave. By the time this beam of "negative light" has traveled a distance L, it has spread out in the lateral dimension by an amount W, given roughly by $W \approx L\Delta\theta \approx L(\lambda/D)$. As the beam spreads out, its amplitude naturally decreases (since it carries a constant amount of energy), until finally it has little effect in cancelling the incident plane wave. Thus the shadow eventually disappears, after going a sufficient distance "downstream". Crudely speaking, we can say the shadow is greatly weakened when the diffraction spreading of the "negative" beam has doubled its width. Then $D \approx W \approx L(\lambda/D)$, i.e.,

$$L\lambda \approx D^2 \tag{50}$$

gives the length L that the shadow extends downstream.

Equation (50) applies to all kinds of waves. You can test it by putting an obstacle in the path of a beam of traveling water waves in a ripple tank. You can perhaps test it with sound waves using a high-pitched whistle as a source and a car as an obstacle. You can test it with light by collecting the light behind the obstacle with a lens that is much wider than the obstacle. If the lens has width D', with $D' \gg D$, then in the focal plane of the lens you have two beams of light focussed into images, the incident plane wave and the

light radiated by the obstacle which superposes with the incident plane wave to form the shadow. The part of the incident plane wave intercepted by the lens has a diffraction angular width $\Delta\theta \approx \lambda/D'$ and will focus to a spot with diameter given by the focal length times this angular width. The "negative light" emitted by the opaque obstacle is focussed by the lens to a spot with diameter given by the focal length times its angular width λ/D, which is much wider than λ/D'. Thus there will be a narrow intense image due to the light collected by the large lens and a broad image due to the "negative light."

5.7 Geometrical optics.

The name "geometrical optics" refers to the study of how light beams behave in optical instruments (which consist of various reflecting and refracting surfaces) in the approximation that one considers only the dominant directions of the beams and does not worry about the spreading of the beams due to diffraction. (The name "physical optics" is sometimes used to describe studies which take into account the wave nature of light and which therefore include interference and diffraction.) The basic "laws" of geometrical optics are the law of specular reflection and Snell's law of refraction. Of course both of these laws are actually due to the wave nature of light; each results from a particular constructive interference.

Specular reflection. Whenever a plane wave is incident on a smooth flat material surface, specular reflection occurs; that is, (a) there is a reflected ray that lies in the plane of incidence (the plane containing the incident ray and the normal to the surface), and (b) the angle of reflection equals the angle of incidence (both angles are measured away from the normal).

Specular reflection is due to constructive interference. The electrons in

the material are driven by the incident wave. They reradiate. The reflected ray for specular reflection is the direction of a constructive interference maximum.

We can most easily understand this by considering our familiar linear antenna array. We let the antenna currents be driven by the electric field of an incident plane wave at non-normal incidence, as shown in Fig. 15.

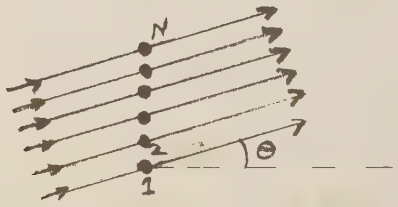


Fig. 5-15. Antenna array driven by plane wave at non-normal incidence.

Now let us examine the distant radiation field due to the antenna currents alone. First consider the central interference maximum. It is easy to see that this occurs in the direction of propagation of the incident beam; antenna No. 1 is excited ahead of No. 2 (in phase) and therefore radiates ahead of No. 2 by the same amount. At a distant point P the radiation from antennas 1 through N will be exactly in phase if the antenna-radiation propagation direction is the same as the incident propagation direction; a certain crest of the wave from No. 1 has to travel farther than that from N, but it started earlier by just the right amount.

It is obvious from the symmetry of the antenna array that the antennas excited as in Fig. 15 will not only form the central interference maximum "to the right" (in the figure) but also a corresponding maximum "to the left." This "image" maximum is the specularly reflected radiation. We see from Fig. 16 that the angle of reflection equals the angle of incidence.

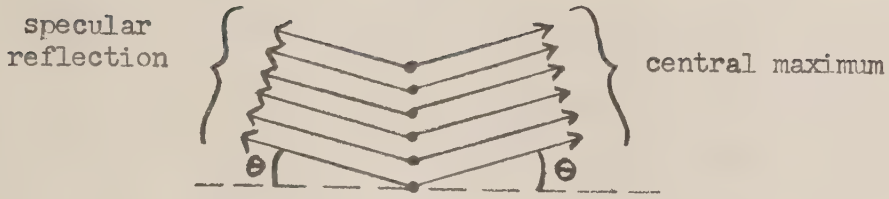


Fig. 5-16. Directions of interference maxima of antennas driven with phase relations of Fig. 15.

Specular reflection from any smooth plane surface is due to constructive interference in a manner completely analogous to that for the closely-spaced antennas.

Image of a point source in a mirror. The surfaces of constant phase for radiation from a point source are spheres. A sufficiently small region on one of these spheres can be approximated by a plane, and we can call the (approximate) plane wave of radiation that passes through this small region a ray. In Fig.17 we see a point source S viewed by means of a mirror. The radiation that enters the lens aperture (pupil) of the eye can be thought of as a "bundle of rays." Two of the rays are drawn in Fig. 17.

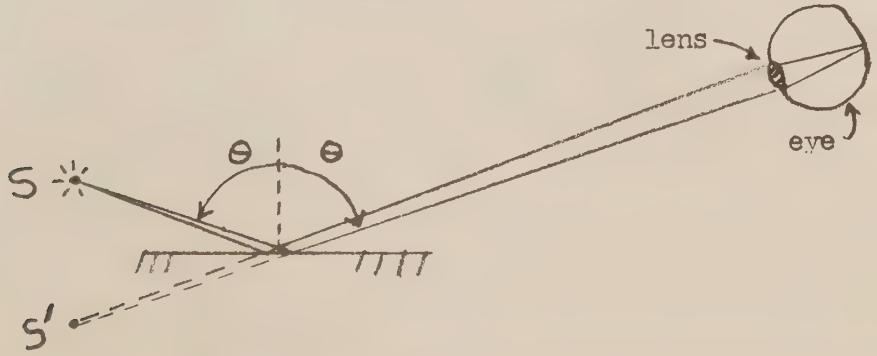


Fig. 5-17. Virtual point image S' of real point source S in plane mirror.

Each ray is specularly reflected from the mirror. The light entering the eye appears to come from a point source S' located "behind the mirror." The source

S' is called a virtual source because there is not a real source of radiation at S'. (The source S is called a real source.)

Refraction. Snell's law. A monochromatic plane wave incident non-normally on a plane boundary surface between two transparent media having different indices of refraction is "bent" as it is transmitted into the new medium. (It is also specularly reflected.) This process is called refraction and is illustrated in Fig. 18.

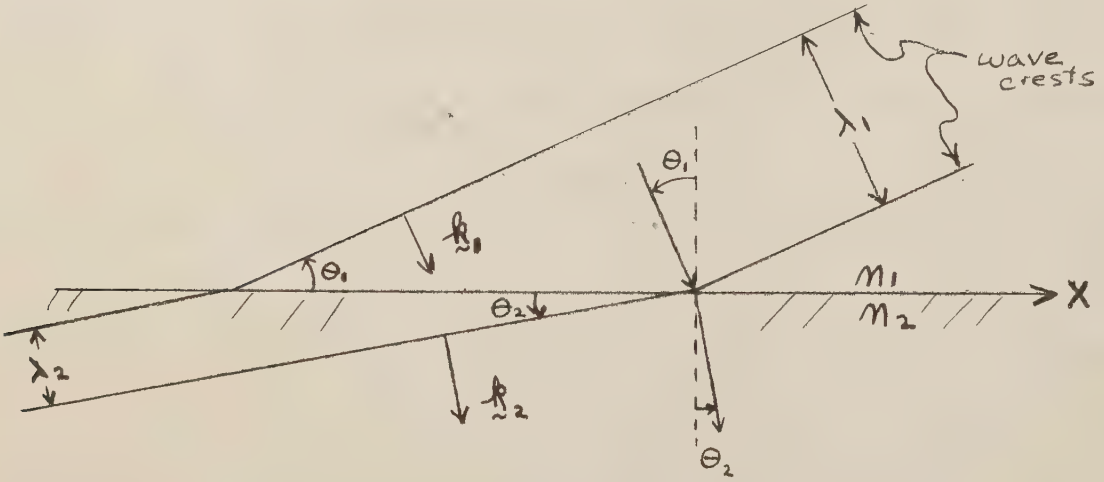


Fig. 5-18. Refraction. This example has $n_2 > n_1$.

It is obvious from the figure that the number of wave crests and troughs per unit distance along $\hat{x}$ is the same on either side of the boundary between the two media, i.e., k_x, the x component of the propagation vector $\underset{\sim}{k}$, is the same in medium 1 as in medium 2. In medium 1, k_x is $k_1 \sin\theta_1$; in medium 2 it is $k_2 \sin\theta_2$. Thus

$$k_1 \sin\theta_1 = k_2 \sin\theta_2 . \tag{51}$$

But k_1 is $n_1 \omega/c$, and k_2 is $n_2 \omega/c$. Thus Eq.(51) gives

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 , \tag{52}$$

which is Snell's law.

Ellipsoidal mirror. In Fig. 19 we see a hollow ellipsoid of revolution with a specularly reflecting inner surface and with a point source of light located at one of the two principal foci, F. From the definition of an ellipse,

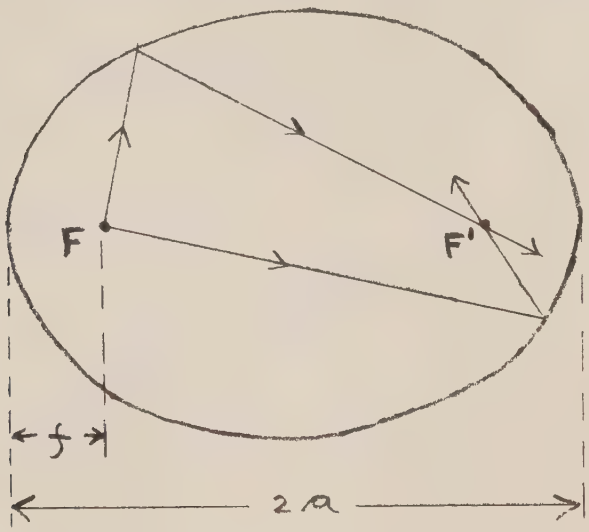


Fig. 5-19. Elliposidal mirror.

the distances from F to the other focus F' are the same for all paths (except the direct one not involving a reflection). Therefore the focus F' is a region of complete constructive interference for radiation emitted by electrons in the surface that are driven by radiation from F. We say that the source at F is imaged at the point F'.

The image at F' is not a point; the phase of the resultant field at a point near F' is within $\pm \pi$ of that at F' provided the point lies within a sphere with radius about $\lambda/4$ and centered at F'. Therefore that is roughly the size of the image at F'.

Parabolic mirror. Imagine that the focal point F and the focal length f of the ellipsoid in Fig. 19 are held fixed, but that the focal point F' is moved to the right; the ellipse is "stretched." If F' is moved infinitely far to the right, the ellipsoid degenerates into a paraboloid. Rays emitted from F then

form a parallel beam (because they still focus at F', infinitely far away). This is shown in Fig. 20.

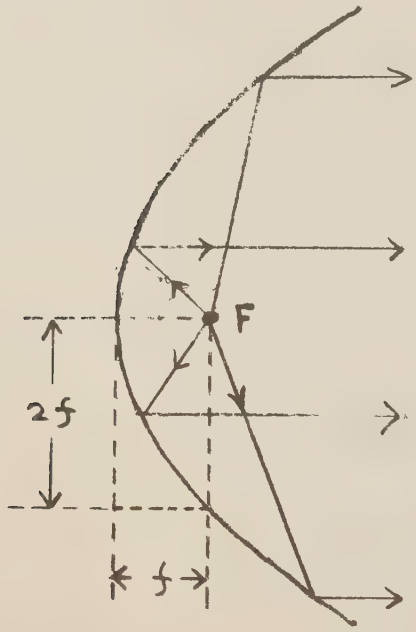


Fig. 5-20. Parabolic mirror.

If the parabolic mirror aperture has a finite diameter D, then a point source at F does not form a perfectly parallel beam. The angular width of the interference maximum is $\Delta\theta \approx \lambda/D$. If D is infinite, we get a perfect plane wave from the point source.

Conversely, an incident plane wave (perfectly well defined in angle) focusses to an image at F that is not a point unless D is infinite. The image has a width $\Delta x \approx f\Delta\theta \approx f\,\lambda/D$.

Spherical mirror. A sphere is said to be "nestled" at the apex of a paraboloid if it is tangent to the paraboloid there and has the same radius as the radius of curvature of the paraboloid there. It is not difficult to show that the radius of such a nestled sphere is 2f. See Fig. 21.

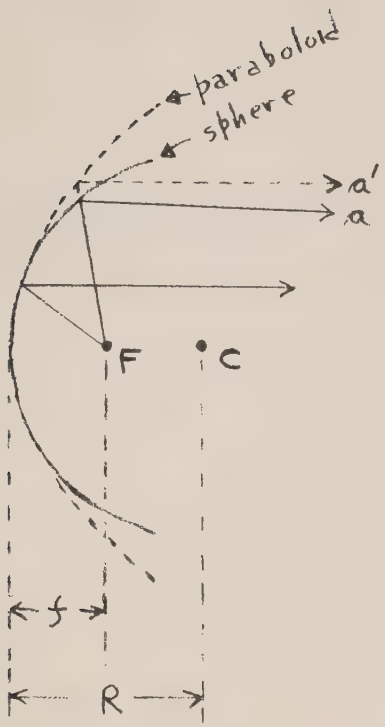


Fig. 5-21. Spherical mirror. Ray a reflected from the sphere is not parallel to the axis, as is ray a' reflected from the paraboloid; this illustrates spherical aberration.

For a small aperture diameter $D \ll 2R$, a spherical mirror is essentially "in contact" with an imagined nestled parabolic mirror. Then a point source at F forms an almost-parallel beam. For large apertures the deviation of the spherical surface from that of a paraboloid produces "spherical aberration." (See Fig. 21.)

Deviation of light ray at near-normal incidence on a thin prism. A "thin" prism is one for which the wedge angle α is so small that we can use the small-angle approximations $\sin\alpha \approx \alpha$, $\cos\alpha \approx 1$. For near-normal incidence, we can also use small-angle approximations for the angle of incidence. Then a mono-

chromatic plane wave at near-normal incidence is deviated "towards the base of the prism" by an angle δ given by

$$\delta = (n-1)\alpha \, . \qquad (53)$$

The deviation δ is a constant, independent of the angle of incidence, as long as we stay near normal incidence. Equation (53) is easily derived as follows (see Fig. 22):

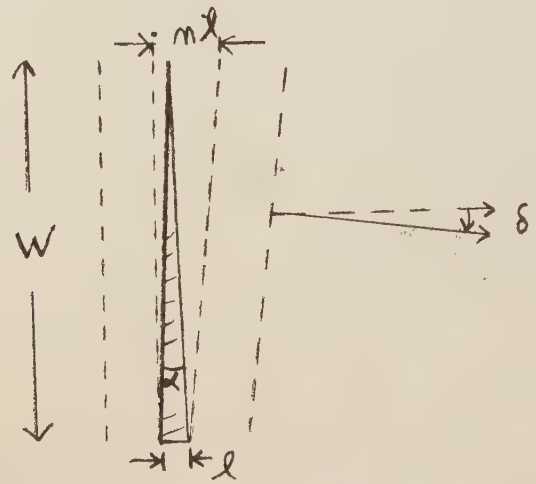


Fig. 5-22. Deviation by a thin prism.

At the base of the prism the wave front transverses distance ℓ at velocity c/n. At the apex, the velocity is n times larger (since the prism thickness is zero there), and the same wave front therefore travels a distance $n\ell$ in the same time. Thus the wave front is ahead by a distance $(n-1)\ell$ at the top. This distance divided by the width W of the prism is (for small angles) the angle of deviation $\delta = (n-1)(\ell/W) = (n-1)\alpha$, which is Eq.(53).

As an example, suppose α is 30° (for which the small angle approximation is still not too bad, for our purpose) and n is 1.50. Then the deviation is 15° according to Eq. (53). That is actually the average deviation, because for glass with an average index of refraction of 1.5, blue light of wavelength 0.45 μ actually has index about 0.01 larger than red light of wavelength 0.65 μ. Therefore the blue light is deviated more than is the red light by about 0.01 α. For α of 30°, blue is deviated about 0.3° more than red. Or in radians, since 30° is about half a radian (1 rad = 57.3°), blue is deviated by about 1/200 radian more than red. On a screen one meter beyond the 30° prism, blue is therefore separated from red by 1/2 cm. A prism spectrometer makes use of this dispersive effect of a glass prism to analyze spectra. In optical instruments involving glass lenses, dispersion leads to chromatic aberration -- i.e., rays of different colors do not focus at the same places. One can still avoid chromatic aberration in a telescope by using a parabolic mirror to gather the light to a focus rather than a refracting lens. (The law of specular reflection holds for all colors.)

Focussing of paraxial light rays by a thin lens. Suppose we have a glass lens in air, with two convex spherical surfaces normal to a common axis of symmetry $\hat{z}$. A light ray is incident from the left, traveling parallel to the axis of symmetry of the lens at distance y = h from the axis. If the lens is "thin", we (by definition) neglect the variation of y as the ray passes through the lens; we also neglect the thickness compared with the focal length. To consider only "paraxial" rays means that we keep h small compared to the radii of curvature of the two surfaces so

that we can use small-angle approximations for all angles of interest.

Let us find the focal point F where an incident ray parallel to the symmetry axis crosses the symmetry axis after deviation by the lens, as shown in Fig. 23.

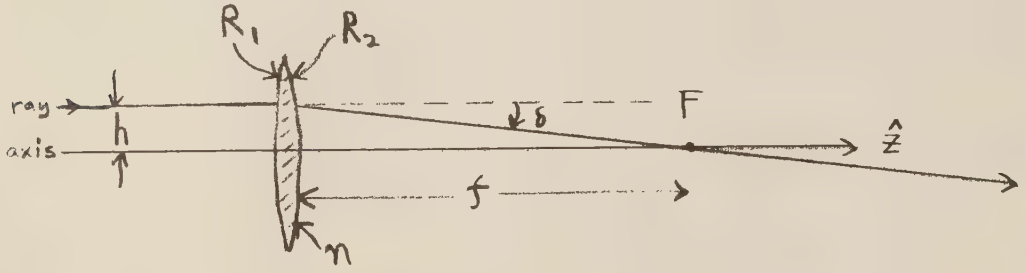


Fig. 5-23. Thin lens. Incident ray parallel to the axis.

We see that if the incident ray is focussed at F it must have been deviated by the small angle

$$\delta = h/f \quad . \qquad (54)$$

Thus we see that the necessary condition for the existence of a common focal point for all parallel incident rays is that the deviation be linearly proportional to the displacement h of the ray from the axis. Thus if Eq. (54) is satisfied for all h (but always assuming small angle deviations), then all parallel rays will be focussed at the same distance f behind the lens. (This condition holds for any similar focussing problem, for example for the focussing of a beam of charged particles by a magnetic lens.)

It remains to be seen whether a thin lens with spherical surfaces satisfies Eq. (54) with f independent of h. This is seen as follows: As far as the ray in Fig. 23 is concerned, it could just as well have been

deviated by an equivalent thin prism. The first surface is at angle h/R_1 to the vertical (where the ray strikes it). The second surface is at an angle h/R_2 to the vertical in the opposite sense. The equivalent prism angle α is therefore $hR_1^{-1} + hR_2^{-1}$. The deviation by the equivalent thin prism is $(n-1)\alpha$, so that we have

$$\delta = (n-1)h\left(R_1^{-1} + R_2^{-1}\right) \quad . \tag{55}$$

We see that Eq. (55) satisfies the condition for a focus, namely that δ be proportional to h; the focal length f is given by [see Eq. (54)]

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \quad . \tag{56}$$

Equation (56) is called the lens-maker's formula.

Now consider a bundle of parallel rays which are not parallel to the symmetry axis but rather make angle θ with the axis. The deviation of a thin prism is independent of the angle of incidence (for small angles). Therefore a ray that strikes the lens at distance h from its center is deviated by $\delta = h/f$, independent of the angle of incidence. That means that any parallel bundle focusses at a point in a plane, called the focal plane, a distance f behind the lens, and the lateral displacement in the plane of the point is $f\,\theta$ from the axis, as shown in Fig. 24.

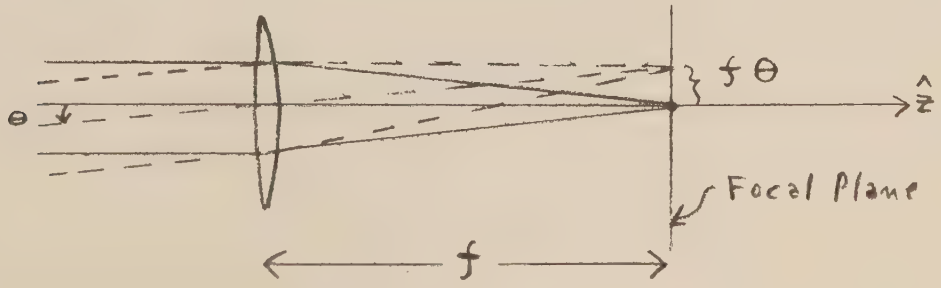


Fig. 5-24. Focal plane.

We have found the point image of a parallel beam, i.e., a beam from an object point (source) infinitely far to the left. Let us now consider an object point $\mathcal{O}$ at distance p to the left of our converging lens and find its image I at distance q to the right. We let $\mathcal{O}$ be on the symmetry axis; then I will be on the axis. Now consider Fig. 25.

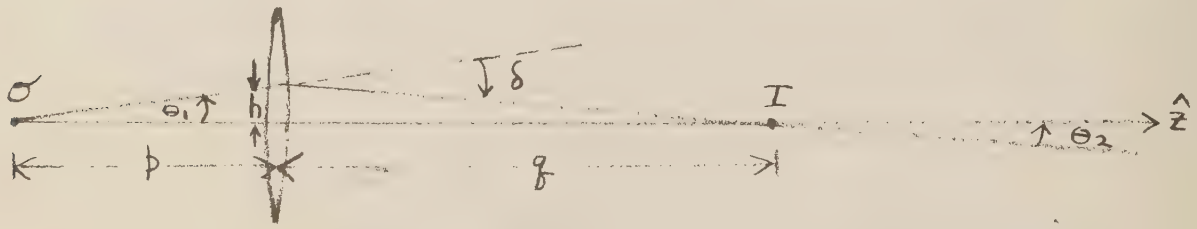


Fig. 5-25. Point image of a point object.

It is obvious from the figure that if we start with a vector pointing from $\mathcal{O}$ in the $+\hat{z}$ direction and then perform the rotations $+\theta_1$, $-\delta$, and $+\theta_2$, we are back to the $+\hat{z}$ axis:

$$\theta_1 - \delta + \theta_2 = 0 \quad . \tag{57}$$

But

$$\theta_1 = h/p \, , \; \theta_2 = h/q \, , \text{ and } \; \delta = h/f \quad .$$

(The deviation is always h/f independent of angle of incidence.) Therefore Eq. (57) gives

$$\frac{h}{f} = \frac{h}{p} + \frac{h}{q} \quad ,$$

i.e.,

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f} \quad . \tag{58}$$

Equation (58) is called the thin lens formula.

If $\mathcal{O}$ is displaced perpendicular to the axis by a small amount $\Delta y_{\mathcal{O}}$, then both rays in Fig. 25 (i.e., the ray both before and after refraction by the lens) are deviated from those shown by the same small angle $\Delta\theta = \Delta y_{\mathcal{O}}/p$. The image point is displaced by an amount

$$\Delta y_I = -q\,\Delta\theta = -q\,\frac{\Delta y_{\mathcal{O}}}{p} \qquad (59)$$

but remains in the same plane, the image plane. The ratio of Δy_I to $\Delta y_{\mathcal{O}}$ is called the lateral magnification, m (lateral):

$$m\ (\text{lateral}) = -\frac{q}{p} \quad . \qquad (60)$$

If $\mathcal{O}$ is displaced along $\hat{z}$ by an amount $dz_{\mathcal{O}}$, the image is displaced in the same direction. Thus if p gets smaller, q gets larger. By differentiating Eq. (58), we find the longitudinal magnification, m (logitudinal):

$$m\ (\text{longitudinal}) \equiv \frac{dq}{dp} = -\frac{q^2}{p^2} = -[m\ (\text{lateral})]^2 \quad . \qquad (61)$$

We have taken p and q positive in our sketches, but they are to be interpreted algebraically. We take p positive when $\mathcal{O}$ is to the left of the lens and negative when $\mathcal{O}$ is to the right; q is positive when I is to the right of the lens and negative when I is to the left. For example, consider Eq. (58). For $p > f$, q is positive. For $p = f$, q is infinitely positive (or negative). For $p < f$, q is negative. That means the image lies to the left of the lens. It is therefore a virtual image. See Fig. 26.

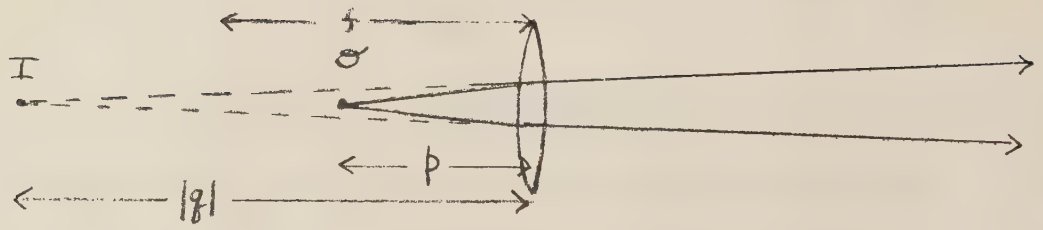


Fig. 5-26. Virtual image. Since p is closer to the lens than distance f, q is negative from the equation $p^{-1} + q^{-1} = f^{-1}$. The image is called virtual because there is no radiation at the image point I.

All of our formulas also apply to diverging lenses when suitably interpreted. In the lens-maker's formula Eq. (56), a concave surface is assigned a negative radius of curvature. That is, R_1 is positive if the center of curvature of the left surface is to the right and negative if it is to the left; R_2 is positive if the center of curvature of the right surface is to the left and negative if it is to the right. Then if the lens is thinner in the center than at the edges, it is a diverging lens with f negative. Then in Eq. (58), with p positive (real object to left of lens), q is negative (image is virtual, i.e., to left of lens).

*Ray tracing. A beam (of light or of charged particles) may pass through a sequence of lenses, slits, prisms (a quadrupole magnet is a "lens" for charged particles; a bending magnet is a "prism") etc., with spaces between the "elements" (lenses, etc.). One may explore the properties of a given system of elements by means of graphical ray tracing. Whenever a ray hits a thin lens (or a "principal plane" of a "thick lens") at distance h from the center, it is deviated by an angle h/f and continues on its way. A graph showing many rays gives a visual picture of the beam. Any lens, thick or thin, may be thought of as a succession of thin lenses with intervening spaces.

If one wants to work algebraically, one can use the thin lens formula and find successive images. But this is usually tedious. An easier method involves the use of column vector rays and matrix operators . A ray may be characterized at any point z by its displacement y from the axis and its slope $\theta = dy/dz$. These may be written as a column vector

$$(\text{ray}) \equiv \begin{pmatrix} y \\ \theta \end{pmatrix} \equiv \begin{pmatrix} y\,(z) \\ \theta\,(z) \end{pmatrix} \quad . \tag{62}$$

The action of a thin lens on a ray may be represented by a linear operation on y and θ : y is replaced by y ; θ is replaced by $\theta - y/f$. In matrix notation this is written

$$\begin{pmatrix} y \\ \theta \end{pmatrix}_{\text{output}} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} y \\ \theta \end{pmatrix}_{\text{input}} \quad . \tag{63}$$

The matrix M that represents a general linear operation on the ray is written

$$\begin{pmatrix} y \\ \theta \end{pmatrix}_{\text{out}} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} y \\ \theta \end{pmatrix}_{\text{in}} \quad , \tag{64}$$

which means

$$\begin{aligned} y_{\text{out}} &= M_{11}\, y_{\text{in}} + M_{12}\, \theta_{\text{in}} \\ \theta_{\text{out}} &= M_{21}\, y_{\text{in}} + M_{22}\, \theta_{\text{in}} \end{aligned} \quad . \tag{65}$$

The matrix that represents an empty space of length L (along $\hat{z}$) is obtained by noting that θ is unchanged and that y is increased by $L\theta$. Thus the matrix equation representing the action of a space L on a ray is

$$\begin{pmatrix} y \\ \theta \end{pmatrix}_{\text{out}} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y \\ \theta \end{pmatrix}_{\text{in}} \quad . \tag{66}$$

Suppose now that the beam goes through thin lens No. 1, followed by space L, followed by thin lens No. 2. The matrix which "takes" any ray

from the input of lens No. 1 to the output of lens No. 2 is

$$\underset{\rightarrow}{M_{1L2}} = M_2 M_L M_1$$

$$= \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_2} & 1 \end{pmatrix} \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{f_1} & 1 \end{pmatrix} . \qquad (67)$$

These are to be multiplied together by the "row-by-column rule" of matrix multiplication to obtain a single matrix that describes the whole system. Ray tracing of charged-particle beams through magnetic lenses and other devices by using a drawing board or matrices (with a fast computer) is common practice in designing particle beams in physics.

Thick lens. Leeuwenhoek's microscope. The world's first microscope was merely a tiny glass sphere. You can easily make one.* Here is how it works. Consider a parallel beam of light incident on the left-hand surface of a glass sphere of index n and radius R. At the first surface, the angle that the surface makes with the vertical for a displacement h from the axis is h/R. A ray at distance h from the axis is therefore deviated by the angle(n-1)h/R. For n = 3/2 (typical of glass),the deviation is then h/2R. This ray therefore would reach the axis inside the glass at focal distance f = 2R. But that is where the back of the sphere is. Therefore this ray

---

* For handy hints on how to make a Leeuwenhoek microscope see "The Amateur Scientist," C. L. Strong, Simon and Schuster, N. Y. (1960). However, do not try making your own glass spheres as they suggest. You may obtain tiny clear glass spheres by the pound from a chemical supply house. Make sure they are clear and not translucent.

hits the axis just as it emerges at the back of the sphere. (This fortunate accident occurs exactly only for n = 1.5.) Now turn the direction of the ray around. That is put a "point bug" at the right-hand surface of a small sphere of radius R ≈ 1 mm. Put your eye close to the left end of the sphere. The sphere focusses rays from the bug into a parallel beam filling the eye pupil. You see the bug comfortably "far away." That is the way the first microscope worked. See Fig. 27.

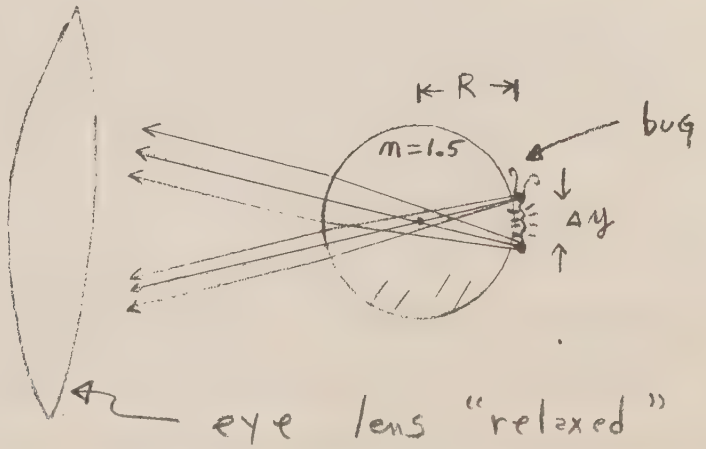


Fig. 5-27. Leeuwenhoek microscope. A point on the bug gives a parallel bundle of rays entering eye. Relaxed eye (i.e., set for distant object) focusses bundle on retina. Separation of retinal images of two points is interpreted by brain (correctly) as an angular size.

The magnification is calculated as follows: Suppose the lateral size of the bug is Δy . Rays from different parts of the bug through the center of the spherical lens are undeviated. Therefore the bug has an "angular size" $\Delta y/R$. That gives the apparent size when viewed with the microscope. Without a microscope, you must hold any small object not much closer than about 25 cm from your eye (otherwise your eye lens rapidly gets tired). At 25 cm, the angular size of the bug would be $\Delta y/25$ cm. The microscope has increased the focusable angular size by a factor (for small angles)

$$m \approx \frac{25 \text{ cm}}{R} \approx 250 \text{ , for } R \approx 1 \text{ mm} .$$

Problems and Home Experiments, Chapter 5

1. How far away is an automobile when you can barely resolve the two headlights with your eye?

2. Venus has a diameter of about 8000 mi. When it is visible as a "morning star" (or "evening star"), it is about as far away as the sun, i.e., about 93 million miles. It looks "larger than a point" to the unaided eye. Are you seeing the true size of Venus?

3. Home Exp. Resolution of the eye. Take two light bulbs of the same power (say 150 watt), one with a clear glass envelope and reasonably small filament ($\approx$ 1 inch by 1/8 inch), the other with a frosted bulb $\approx$ 3 inches in diameter. Find out by experiment how far away you must walk before the two lights have the same apparent size. (It will be a block or two.) At this same large distance compare the apparent sizes of two frosted bulbs having the same actual size but differing in power by a factor of two or three. How do you explain the result?

*4. Home Exp. Diffraction grating Moire pattern. You need a white "line" source and two identical gratings. [The best line source (which you will need for many of these experiments) is a "display lamp." For example, one of about 40 watts with a three inch long straight filament in a clear glass envelope is available for about 40 cents in grocery and hardware stores). Your optics kit only has one grating. More gratings are available for about 25 cents each (or about 10 cents each in boxes of 100) from, e.g., Edmund Scientific Company, Barrington, N. J. 08007.] With the line source oriented vertically, look through one grating (hold it up close to one eye) and orient

the grating so that the colors are spread out horizontally. Now superpose the second grating on the first. Carefully rotate it so as to superpose exactly the first order images from the two gratings. With care you will succeed (in a minute) in obtaining "black stripes" across the colored first order image. Here is the explanation. The line spacing on the grating is d. Suppose the space between the planes of the two gratings is s. Think of two picket fences superposed with a slight space between them or of two identical screens parallel to one another. At some angles the grating scratches will lie one behind the other. At other angles the scratches of one grating will lie (in projection) half way between the scratches of the other. At these angles the effective number of scratches per unit length i.e., d^{-1}) is doubled. Now comes the physics: Why do you get the black stripes? Do they correspond to angles where the effective number of lines is "single" or "double?" Given the number of lines per cm, d^{-1}, for each grating, how can you determine the spacing s? Given s, how can you determine d?

5. Home Exp. Silk-stocking diffraction pattern. You need a sheer silk (or nylon) stocking and a point source of white light. A street light will do for the point source. [The best point source for this and other experiments is made from a 6 volt flashlight, for example a "camper" flashlight, with a bulb that has a filament about $\frac{1}{2}$ mm long. To get a good point source, remove the glass lens and cover the parabolic reflector with a piece of dark cloth or paper (with a hole cut for the bulb). Or simply look at the bulb from the side, out of the beam from the reflector.] Look through the stocking at the point source. From the pattern you see you could determine the average

thread spacing and the number of sets of threads at different angles. Fold over many layers and look at the source again. The pattern of concentric circles you see is similar to an x-ray "powder diffraction pattern."

6. Home Exp. Long-playing diffraction grating. Look at a white point source reflected at near-glancing incidence on a 33 rpm record. The record grooves make a good reflection grating. Measure crudely the wavelength of red and of green light using the record. Describe your method. How can you easily determine the location of the zeroth order "specular" maximum?

7. Home Exp. Which side has the scratches? One side of the plastic of your diffraction grating is smooth. The other side has the scratches. You can find out which side has the scratches by looking through it at a white source after rubbing one side of the grating with an oily finger; then clean it and try the other side. What is the explanation?

8. Consider nestled spherical and parabolic mirrors as shown in Fig. 21. Take the $\hat{z}$ direction to the right (along the axis of symmetry) and $\hat{x}$ transverse to $\hat{z}$; take $x = z = 0$ at the apex of the mirrors.

(a) Show that the parabolic surface is given by

$$z = \frac{x^2}{4f} .$$

(b) Show that the spherical surface is given (for $x \ll f$) by

$$z = \frac{x^2}{4f} + \frac{x^4}{64f^3} + \cdots .$$

(c) Compare a spherical mirror with aperture of diameter D and with focal length f to a parabolic mirror with the same D and f .

For the spherical mirror, consider the angular deviation $\delta\theta$ of the "worst" rays (near the rim of the aperture) due to spherical aberration. ($\delta\theta$ is the deviation from the $\hat{z}$ direction for rays from a point source). Show that $\delta\theta$ is less than the diffraction angular width $\Delta\theta \approx \lambda/D$ provided that

$$D < 4f\left(\frac{\lambda}{4f}\right)^{1/4} .$$

Thus (for example) for visible light and for focal length $f \approx 50$ inches, a spherical mirror is about as good as a parabolic mirror provided the mirror diameter D is less than ~ 3.5 inches.

9. A plane slab of glass of thickness t and index n is inserted between an observer's eye and a point source. Show that the point source appears to be displaced to a point closer to the observer by an amount $[(n-1)/n]t$.

10. A "corner reflector" consists of three plane mirrors joined so as to form an inside corner of a rectangular box. Show that a light beam that strikes a corner reflector is directed back at 180° to its original direction, independent of the angle of incidence, as long as it hits all three surfaces.

11. Show that a plane wave <u>normally</u> incident on one face of a wedge-shaped prism of angle A is deviated by an amount θ_{dev} where

$$n \sin A = \sin (A + \theta_{dev}) .$$

*12. Find the matrix that represents the combined behavior of (from left to right) thin lens No. 1 with focal length $f_1 = +100$ cm, a space $L = 40$ cm, and thin lens No. 2 with $f_2 = -50$ cm (diverging) for rays traveling left

to right and for rays right to left. Find the right-hand focal point (the location of the focal plane for a ray parallel to the axis incident from the left) and also the left-hand focal point (the location of the focal plane for a parallel ray incident from the right). Be careful with sign conventions on the right-to-left rays. Sketch the path of rays through the system for parallel incident rays from each side and for rays from an object 200 cm to the left of lens No. 1.

13. Home Exp. Interference in thin films. The colors in soap bubbles and in oil slicks are due to interference between light reflected from the front surface and light reflected from the back surface of the film. The amplitude reflection coefficient has opposite sign for light incident from air to oil than from oil to air, so that at near-zero thickness the two reflected waves cancel one another to zero. (If they didn't, there would be a discontinuity in the reflected flux at zero thickness, since at zero thickness there has to be zero reflection.) If the film is a quarter-wave thick (wavelength in the film, not the vacuum wavelength) then "down and back" is $\frac{1}{2}\lambda$, giving 180° relative phase shift between contributions from the two surfaces and consequently giving completely constructive interference (they were already 180° out of phase at zero thickness). As the film thickness varies you get alternations of constructive and destructive interference at values of thickness that depend on the wavelength.

Here is the experiment. Fill a pan with water. Put one drop of oil on the water and wait for the oil to spread. Look at the sky (or any other broad source of light)reflected in the oil film. [It helps to use hot water (the oil spreads more easily) and to use a light penetrating oil, e.g.,

3-in-one oil. Salad oil is too heavy; it doesn't spread over the surface. A black cloth or paper on the bottom of the pan helps; it gives a dark background so that the reflection from the surface is not overpowered by reflection from the bottom of the pan.] Notice that no colored fringes appear until the film has spread to 10 cm by 10 cm or so. Why is that? Watch the colored fringes as the film continues to spread. When it gets thin enough, you will get no more fringes; it will be "black" where the oil is thinnest. That is the region where the film is less than a quarter-wave thick. Use this fact to estimate crudely the wavelength of visible light. Take the "black" region to be (for a crude estimate) about an eighth of a wavelength thick; estimate the area of the film, and use that and the original size of the drop to find the film thickness, which then gives the wavelength.

Another easy method to obtain thin-film interference fringes is to press two microscope slides together so that they are in intimate contact. If you buy a box of a dozen slides (they cost about 35 cents/dozen), you will usually find many of the slides stuck together and giving nice fringes.

14. Home Exp. Single-slit diffraction pattern. Use a line source (see Home Exp. 4). Look at the source through a single slit formed by your two parallel thumbnails held close to your eye. A better single slit is made by cutting a slot in the edge of a piece of cardboard with a scissors. The slot can be two or three mm wide. Hold it in front of your eye and turn the cardboard so that the "apparent" (i.e., projected) width of the slot can be made as narrow as you wish. Vary the apparent slot width and see the effect on the width of the pattern. A better slit is obtained by sticking two

pieces of opaque scotch tape on a microscope slide with their edges parallel, so as to give a slit perhaps 1 mm wide. Hold the slide close to your eye with the tape side towards the light source (the fringes are slightly clearer that way than when the tape side is nearest your eye) and vary the apparent width by tilting the slide. To see the effect of wavelength, cover the source with red or blue cellophane. Also, look at a point source as well as a line source.

15. Home Exp. Diffraction at a single straight edge. This experiment is done using the same materials as the previous experiment but merely looking at the line (or point) source just past a single edge held near your eye. The pattern resembles "half of a single-slit pattern." The light "bends around the corner into the shadow" behind the straight edge. Why then is the apparent position of the interference fringes out in the bright region, rather than in the shadowed region? (Make a sketch including source, edge, and eye; then the answer is obvious.) This experiment can also be done using a point source. A good point source (besides that mentioned in Home Exp. 5) is the reflection of the sun from a curved surface of chrome or from the windshield of an automobile one hundred feet or so away. Your thumbnail makes a good enough straight edge.

16. Home Exp. Double-slit pattern. The easiest way to make a double slit is with a microscope slide and opaque scotch tape. A good slit width is about $\frac{1}{2}$ mm, which you can easily manage by laying one piece of tape carefully beside another. To study the effect of the slit separation d, let the opaque tape between the slits be tapered from a maximum of about 2 mm to a

point. Thus if each slit is $\frac{1}{2}$ mm wide, the slit separation varies from $2\frac{1}{2}$ mm to $\frac{1}{2}$mm. Now tilt the slide so that the projected dimensions are as small as you please. While looking at the line (or point) source, cover and uncover one of the slits with a piece of paper. Then you can tell what part of the pattern is the "single-slit" pattern due to the width of each slit and what part is the "double-slit" pattern. Here is a problem: show that the number of "two-slit" interference maxima within the central "single-slit" modulation is given approximately by $2d/D$, where D is the width of one slit. Another good way to make slits is to tape a piece of aluminum foil on a microscope slide and cut the foil with a razor blade and straight edge.

17. Home Exp. Coherence; size of a "point" source. Use a single-slit of known (estimated) width. Use a red cellophane filter over the source. Stand ten feet or so from the source so that you obtain a sharp single slit pattern. Now move closer to the source. Find the distance L at which the single-slit pattern "washes out." [It washes out because different parts of the filament of the flashlight bulb (if that is your point source) are independent light sources, as discussed in Sec. 5.4.] Use your estimate of the sizes of the source and the slit and your measurement of the distance L at which the pattern washes out to estimate the wavelength of the light, using the relation derived in Sec. 5.4, $d(\text{source})\, D(\text{slit}) \approx L\,\lambda$.

18. Home Exp. Coherence. Lloyd's mirror, the "guaranteed coherent double slit." If you hold an ordinary double slit in front of your eye and look at a broad source like the sky or a frosted light bulb, you will see no interference pattern. We shall now design a double-slit that will give a two-slit interference pattern even when looking at a frosted light bulb. For this experiment you need as narrow a single slit as you can make (with the crude

techniques we suggest). The slit should be about 0.1 mm wide. You can barely achieve this by laying one piece of opaque scotch tape parallel to another. (A good alternative method is to tape a piece of aluminum foil onto a microscope slide so that the foil is stretched flat and taut against the glass. Cut the foil gently with a razor blade against a straight edge.) Now take a second slide and place it edgewise to the first slide (the one with the slit)and parallel to the slit, so that the mirror image of the slit in the second slide is parallel to the first slit. Stick the second slide to the first with a big glob of putty or modeling clay (non-drying putty, e.g., Nu-Glaze glazing compound works well), so that you can easily wiggle the second slide to adjust it, yet so that it will stay in place when you don't push on it. Adjust the mirror (the second slide) so as to get as narrow a separation between the slit and its "image slit" as you can manage -- say $\frac{1}{2}$ mm. Do this by holding the assembly a foot or so from your head so that you can easily focus your eyes on the "double" slit while you hold it in front of a bright background and adjust the mirror. When you have a good double slit, bring the assembly up close to one eye and focus the eye at a large distance (i.e., on the light source). Look for three or four "black streaks" parallel to the "coherent double slit." These are interference zeros due to destructive interference between the light coming from the real slit and that coming from the image slit. The image slit is of course always completely coherent with the real slit (why?). Because of the phase reversal in reflection, the slit currents and the "image slit currents" are 180° out of phase. Therefore the fringe at the plane of the mirror is "black" -- an interference zero. Here is a question, to be answered both by experiment and by "theory": Are the "bright" lines between the "black"

lines exactly as bright as the bright background that one sees with just the single slit? Brighter? Dimmer?

19. Home Exp. Spectral lines. Pour some table salt on a spoon (one that you don't mind ruining), wet the salt so it will stick to the spoon, and set the spoon in the flame of a gas stove. Look at the yellow flame through your diffraction grating (this is easiest at night in a darkened room). Notice that the first (and higher) order images of the yellow sodium flame are as sharp and clear as the zeroth order image. That is because the yellow light is a "spectral line" having narrow bandwidth. (Actually the yellow light from sodium is a "doublet" of two lines with wavelengths 5890 and 5896 A^{o}.) Now look at a candle. In zeroth order, it does not look terribly different from the sodium flame. They are both yellow. But in the first order diffraction image, the candle is very much spread out in color, whereas the sodium remains sharp.

Here are other convenient sources of sharp spectral lines; look at them through your grating: Mercury vapor: fluorescent lamps, mercury-vapor street lights, sunlamps. (A sunlamp is convenient in that it screws directly into an ordinary 110 V AC socket. It is probably the cheapest source of mercury-vapor spectral lines. Cost is about $10.) Neon: many advertising signs. Neon has a profusion of lines; you see "many signs." A cheap source is a G.E. bulb NE-30 which screws directly into a 110 V AC socket (the cost is about $1). Another is a "circuit continuity tester" which plugs into any wall receptacle and costs about $1 (at a hardware store). Strontium: strontium chloride salt (available at a chemical supply house for about 25 cents/oz.); dissolve a little in a few drops of water and put it in the

gas flame on your ruined spoon. The wavelength of the red line is a famous length standard. Copper: copper sulfate; availability and technique as for strontium chloride. Gives beautiful green color. Hydrocarbon: Look at your gas flame in the first order spectrum. It is sharp and clear. The blue color of the flame is therefore due to one or more narrow lines.

20. Home Exp. Mailing tube spectrometer. Fraunhofer lines. Use a mailing tube $1\frac{1}{2}$ to 2 ft. long. Mount your diffraction grating at one end. Mount a single slit on the other end. The slit is best made with two single-edge razor blades. Glue one blade permanently in place. Stick the other on with non-hardening putty (glazing compound, obtainable in any hardware store) so that you can easily adjust it (narrower for better resolution, wider for more light). Look at the spectra mentioned in the previous experiment. Problem: should you be able to resolve the sodium doublet (wavelengths 5890 and 5896 A^{o})with this spectrometer? (Answer: No -- the line separation given by this grating is just about equal to the image width due to diffraction in your eye pupil.) Can you resolve it by using a longer mailing tube? (Answer: No. There are two ways to improve the resolution. One is to get a grating with smaller d. The other is to increase the number of lines that are used, i.e., to increase the width D of grating used. With the design above, D is the width of your pupil, about 2 mm. If you add a telescope with an objective lens of diameter 2 cm, and if all the rays that enter the objective lens get through the pupil of your eye, then with the diffraction grating at the objective lens, your angular resolution, λ/D, is improved tenfold.)

With this simple spectrometer you can see the Fraunhofer lines in the spectrum of the sun. Go outside on a sunny day. Lay a pile of half a dozen sheets of white paper on the ground (more than one so that it is as "white as possible"). Look at the sunlit paper with your spectrometer. Use a coat or blanket to cover your head to keep out stray light; otherwise you will have difficulty seeing the first order spectrum. Also, use the edge of the tube to "hide" the blindingly bright zeroth order light. Adjust the slit to about $\frac{1}{2}$ mm. Look for three or four or five dark lines crossing the continuous spectrum of the sun. If you don't see anything, keep trying -- adjust the slit width for comfortable intensity. Another technique is to cover the slit with several layers of waxed paper, use a very narrow slit, and look at the sky near the sun, varying the intensity by how close you come to pointing the spectrometer at the sun. The dark Fraunhofer lines you see are absorption lines. Atoms in the relatively cool outer gas mantle of the sun are driven by the continuous spectrum emitted by the hot sun. Those frequencies that correspond to natural resonances of the atoms excite the atoms. This takes energy out of the continuous spectrum at the resonant frequency. The outer gas is actually opaque at those frequencies, so that the spectrum has corresponding "black lines" at colors where the sunlight has been completely absorbed. The easiest lines to see are some closely spaced lines in the yellow-green due to iron, calcium, and magnesium; the H line in the blue-green due to hydrogen; and several closely spaced lines in the blue due to hydrocarbons -- similar to the emission lines you see with a gas flame. The sodium D line is also present, but slightly harder to see (for me at least). To see where to look for it, look at the sodium emission line by throwing salt on a gas flame. That is the color that is "missing" in the Fraunhofer spectrum.

21. Home Exp. Ripple tank. Put a dark cloth on the bottom of a pan of water and extend the cloth up the sides to form a "beach." The beach absorbs traveling waves in the water without much reflection. The dark cloth gets rid of light reflections from the bottom of the pan. Illuminate from above. Use a broad source, such as the sky -- but look at the waves from angles where you are looking at the reflection of something dark when the surface is smooth. The waves then reflect bright sky where formerly it was dark. For a source, use a hacksaw blade (25 cents at any hardware store) clamped to a table. Screw a small piece of wood on the end of the blade. (The added mass provides energy storage so as to keep the motion going longer.) Into the piece of wood screw two long round-head screws. Adjust until the screw heads just touch the water. Vary the length of the blade and the spacing between the screws so as to get nice interference minima in the two-source interference pattern. The main difficulty in seeing them is the illumination. You will have to optimize this by varying the illumination and the position of your head.

A very striking variation of the two-source pattern is provided by driving each of the two screws independently with a separate hacksaw blade. Adjust the frequencies so that they are almost identical. Then the interference pattern rotates with maxima and minima alternating at the beat frequency.

Using a single driving source (one screw), put an obstacle in the path of the waves. Estimate the distance downstream at which the "shadow" has washed out. How does it compare with the estimate $L \approx D^2/\lambda$ obtained in Sec. 5.6?

22. Simple magnifier. You can hold a small object at about 25 cm from your unaided eye and examine it without fatigue. If its height is h (in cm) it subtends an angle $h/25$ (radians) at your eye, and this determines the size of the image on the retina. Now use a lens of focal length f (cm). Hold the lens just in front of the eye. Bring the object in closer. When the object is at the focal point of the lens, each point on the object will give a parallel bundle of rays out of the lens and into your eye. This is easy for you to focus -- your eye lens is relaxed. Show that the angular size of the object increases by a factor $25/f$ (assuming small angles so that you can use small-angle approximations).

23. Telescope. A telescope consists of two lenses. The first is the "objective" lens, which forms a real image of a distant object. To a good approximation,the image is in the focal plane of the objective lens. Show that the height of this image is $h_1 = f_1 \theta_o$, where f_1 is the focal length of the objective lens and θ_o is the angular size of the distant object. The second lens of a telescope is called the "eyepiece." It is effectively a simple magnifier that is used to examine the real image formed by the objective lens. Show that if the eyepiece is adjusted so that the image formed by the objective lens is in the focal plane of the eyepiece, the overall angular magnification of the telescope is f_1/f_2. Show that not all of the rays from a distant object that strike a very large diameter objective lens get into your eye and that in fact the "useful diameter" of the objective lens is about f_1/f_2 times the diameter of your eye pupil. Thus in an eight power telescope, if the exit beam is a parallel beam of width 4 mm (twice as wide as your eye pupil, so that your eye need not be

perfectly lined up), the objective lens should be 32 mm in diameter, but a larger diameter is a waste of objective lens.

*24. Telescope design. Read Home Exp. 20 (Mailing tube spectrometer). Design a telescope to use with your mailing tube spectrometer such that you can resolve the sodium doublet. The design should specify the entire spectrometer, i.e., slit, length of tube from slit to objective lens, diameter of objective lens (the grating is 1 inch square and should go at the objective), focal length of objective lens, focal length of eyepiece lens, and diameter of eyepiece lens. Look up your required lenses in (for example) the Edmund Scientific Company catalog (obtainable free from Edmund Sci. Co., Barrington, N. J. 08007) to see if they are available.

25. Microscope. A microscope is like a telescope in that there are an objective lens to form a real image of the object and an eyepiece to examine this image. The bug to be examined is nearly at the focal plane of the objective. The image is formed a long way from the objective -- say 20 cm. This distance is essentially the length of the barrel of the microscope. Show that if the eyepiece is adjusted so that the image formed by the objective is in the focal plane of the eyepiece, the magnification is $25\ L/f_1 f_2$, where L is approximately the barrel length (distance from objective lens to first image), f_1 and f_2 are the focal lengths of objective and eyepiece, and the magnification is the ratio by which the apparent size of the bug is increased compared with holding it 25 cm from your eye.

26. How wide is a "plane wave" from a distant point source? It is often said that the traveling wave from a distant point source is like a plane

wave over a "limited region" transverse to the line of sight from the point source to the field point. How limited is the region? Suppose the source is at distance L and we wish to consider a circular plane region of radius R transverse to the line of sight from the source. How large can R be so that the phase at the center of the circle and at the edge of the circle differ by less than $\Delta\varphi$ radians? Answer: The phase at the center of the circle is ahead of that at the edge (the center is closer to the source) by an amount

$$\Delta\phi = \pi R^2/L\lambda \quad .$$

Thus the phase is "the same" over the entire plane of the circle to the extent that the area of the circle is small compared with $L\lambda$.

27. The world's largest parabolic radio antenna (at present) is at the National Radio Astronomy Observatory at Green Bank, West Virginia. It is a paraboloid dish 300 feet in diameter. What is its angular resolution in radians and in minutes of arc (the units used by astronomers) for the famous 21 centimeter radiation of hydrogen? Answer: A point source will look like a volley ball at a distance of 300 feet.

28. Home Exp. Color standard. It is good for you to know the approximate wav lengths of colors that you recognize. A color plate of a spectrum helps(we hope have one in the final edition). Meanwhile, use your Edmund grating, a point or line source, and a tape measure to construct a reference system. For example, with the source near a wall, stand 10 ft. or so from the wall at a standard position. Put labeled markers at 500 A^{o} intervals along the line on the wall where the colors appear to fall. Calculate the position of the markers from the geometry and the known line spacing of the grating.

29. Home Exp. Pitch standard. Some of our home experiments have involved using a TV set as a stroboscope. The screen turns on and off at 120 cps. To see a hacksaw blade "stand still" at 120 cps, it is helpful to have a rapid method to tune the blade to 120 cps. The easiest method is to use the pitch of the sound of the vibrating blade. For a comparison pitch at 120 cps, one can use any one of many common household devices containing a transformer which gives a "60 cycle hum," which is actually a hum at 120 cps. (Question: Why is the hum at 120 cps when the voltage varies at 60 cps?) Common sources of 120 cps hum are (1) a radio just after it is turned on and is warming up (after it is warmed up you can still hear the hum if you are not tuned in to a station); (2) a fluorescent lamp; (3) an electric clock (put your ear right against the clock); (4) a "high intensity" lamp that uses a 6 volt or 12 volt auto headlamp bulb.

30. Home Exp. Fluorescent lamp stroboscope. You can use a fluorescent lamp or a sunlamp as a stroboscope at 120 cps. (It is not as good as the TV because it does not turn completely off between the times when it is bright.) Here is an interesting experiment that requires a fluorescent lamp and a TV set. The experiment is to strobe something which vibrates close to 120 cps with both the fluorescent lamp and with the TV (which flashes at 60 cps) and see whether the beat frequencies are identical in the two cases. One technique is to use a hacksaw blade tuned to 120 cps. Compare the beat rate in the two cases. Another technique is to use a toy gyroscope. As the gyro slows down, the spokes "stand still" at various times. Arrange things so that you can illuminate the gyro with the fluorescent light and simultaneously look through the gyro at the TV screen. Then you can see whether the beats occur simultaneously. After you have found the answer, explain it. (The 120 cps from

the fluorescent lamp ultimately depends on the frequency of rotation of a big flywheel at some power station. What does the 60 cps from the TV depend on? Is it generated in the TV set, using the variation of line voltage from your wall receptacle? Is it generated at the TV station, using the 120 cps from their line voltage? Does it make sense for the frequency to suffer a slight "loss in transmission" from the power station to your home or from the TV broadcasting antenna to your home? Is it possible that the power station that delivers power to the TV station corresponds to a different flywheel than that which delivers power to your home and that these two flywheels differ in frequency by perhaps $\frac{1}{2}$ cps? What effect would that have on synchronous electric clocks served by different flywheels after an hour or so?)

Advanced Topics

1. "Microscopic" examples of weakly-coupled identical oscillators.

First, read Sec. 1.5 on weakly-coupled identical pendulums and then the paragraph "Esoteric examples"(page 42,Chap. 1), which introduces this topic.

Here are some examples of weakly-coupled oscillators taken from atomic physics and elementary-particle physics. In each case there are "two identical degrees of freedom" which are "weakly coupled," so that there are two "normal modes" with frequencies ω_1 and ω_2. However, we are not dealing here with macroscopic mechanical systems. Newton's laws do not suffice. The understanding of these "microscopic" systems requires quantum mechanics. Nevertheless, there is a great mathematical similarity in the behavior of the microscopic systems we shall describe to that of two weakly-coupled pendulums. The physical interpretation is very different, however. For the coupled pendulums, the square of the amplitude of a pendulum is proportional to the energy (kinetic plus potential) of that pendulum. The energy "flows" back and forth from one pendulum to the other at the beat frequency. For a system described by quantum mechanics, the square of the amplitude describing a system of one degree of freedom (actually the absolute square -- the amplitudes are always complex quantities in quantum mechanics) gives the probability that that degree of freedom is "excited" (i.e., has all the energy). The probability "flows" back and forth from one degree of freedom to the other at the beat frequency, ω_1 and ω_2. The energy is "quantized;" it cannot "subdivide" to "flow." In the case of the pendulums, the total energy of both pendulums

is constant. In the microscopic systems, the total probability that one or the other degree of freedom is excited is constant (the probability is unity as long as the system does not somehow lose the excitation energy.) Here are two famous examples. (You will encounter them when you study quantum mechanics.)

Ammonia molecule. This molecule is NH_3 (one nitrogen and three hydrogen atoms). (See Vol. 2, page 316.) The three H atoms form an equilateral triangle. Call the plane of this triangle the H_3 plane. The N atom has two possible positions where it may vibrate. (These correspond to pendulums a and b.) One (position a) is on one side of the H_3 plane; the other (position b) is on the other side. The N atom cannot easily get from a to b, or vice versa, because there is a potential "hill" or barrier between a and b. Classically (i.e., according to Newtonian mechanics, without quantum mechanics), a and b are positions of stable equilibrium, and the N atom vibrating at a can never get over to b. (In the pendulum analogy, it is as if we removed the coupling spring. Then if pendulum a is vibrating and b is at rest, that condition will be preserved forever, neglecting friction.) But quantum mechanics introduces a "coupling" between a and b in that it allows "potential-barrier penetration." Suppose the molecule starts at time $t = 0$ in the quantum-mechanical state for which N is definitely at a. Then the probabilities are given by $|\psi_a|^2 = 1$, $|\psi_b|^2 = 0$ (i.e., unit probability for N to be vibrating at position a; zero probability for it to be at position b). However, one finds that this condition is not maintained. In fact, one finds (by

solving the Schrödinger equation) that for this starting condition, the probability of finding N at a, namely $|\psi_a|^2$, and that to find it at b, namely $|\psi_b|^2$, satisfy

$$|\psi_a|^2 = 1/2\,[1 + \cos(\omega_1-\omega_2)t] \tag{1a}$$

$$|\psi_b|^2 = 1/2\,[1 - \cos(\omega_1-\omega_2)t] \tag{1b}$$

Equations (1) are remarkably similar to Equations (72), Sec. 1.5. The total probability that N is at one place or the other is of course unity, as we find by adding Eqs. (1a) and (1b).

Just as with coupled pendulums, an ammonia molecule can be "started up" so that it is in one normal mode (or the other). It turns out that if the molecule is in the mode with slightly higher frequency (call this ω_2; $\omega_2 > \omega_1$) then it is unstable. It likes to emit electromagnetic radiation and get into mode ω_1, which is called "the ground state." This radiation can be detected. Its frequency is the beat frequency, $\nu_2 - \nu_1$, which has the value

$$\nu_{beat} = \nu_2 - \nu_1 \approx 2 \times 10^{10} \text{ cps.}$$

This corresponds to a wavelength ($\lambda = c/\nu_{beat}$) of about 1.5 cm, which is in the typical "radar" or "microwave" range. If one sends a microwave beam of frequency 2×10^{10} cps through ammonia gas, some of the microwave photons induce transitions from the ground state (mode 1) to the excited state (mode 2). Thus energy is exchanged between the microwave beam and the gas by exciting the molecules. Similarly, an excited molecule may "decay" to its ground state, adding one photon to the microwave beam. This energy

exchange between the microwave beam and the ammonia gas provides the basis for the "ammonia clock." Absorption of energy from the beam "winds up" the clock. The "probability flow" from state a to b and back at the beat frequency provides the "escapement mechanism" of the clock. This clock and its descendants provide the world's most accurate clocks.

Neutral K Mesons.

Another example of a fascinating system with a behavior analogous to that of weakly-coupled pendulums is that of the neutral K mesons, which are "strange particles." They are very strange and are not yet understood completely. This system has two degrees of freedom, called the K^o meson and the $\overline{K}^o$ meson. They are "coupled" because either of them can interact with two pi mesons or pions (among other things) via a "weak interaction." (The pions are analogous to the spring.) There are therefore two normal modes, which are called the K_1^o meson and the K_2^o meson. Unlike the modes we have previously discussed, one of these modes (the K_1^o mode) is strongly damped. The other is weakly damped. If the system starts out at $t = 0$ with unit probability of being in the K_1^o mode, that probability decreases exponentially with time and is given by e^{-t/τ_1}. A similar (but much smaller) damping occurs for the K_2 mode. The "loss or probability" corresponding to the damping is the result of radioactive decay of the modes into other particles. For example, K_1^o decays mostly into two pions and τ_1 is the mean decay time for K_1^o.

If the system starts at t = 0 with unit probability of being in the K^o state (call that state a), and if there were no damping, then the probability at a later time of the system's being in that same state (K^o) would be given by Eq. (1a). The corresponding probability of finding the system in state b ($\overline{K}_o$) would be given by Eq. (1b). Because of the damping, these formulas must be modified to

$$|\psi(K^o)|^2 = 1/4\ [e^{-t/\tau_1} + e^{-t/\tau_2} + 2e^{-1/2(t/\tau_1 + t/\tau_2)} \cos(\omega_1-\omega_2)t] \ , \quad (2a)$$

$$|\psi(\overline{K}^o)|^2 = 1/4\ [e^{-t/\tau_1} + e^{-t/\tau_2} - 2e^{-1/2(t/\tau_1 + t/\tau_2)} \cos(\omega_1-\omega_2)t] \ . \quad (2b)$$

Note that when $\tau_1 = \tau_2 = \infty$ (no damping), Eqs. (2) are identical with Eqs. (1).

It is an interesting exercise to devise a damping mechanism that will damp only mode 1 and another that will damp only mode 2 for the weakly-coupled pendulums. Then the equations for the energy in the pendulums will be like Eqs. (2) instead of Eqs. (1).

2. Dispersion relation for de Broglie waves.

A de Broglie wave to describe a single particle of definite energy has the form

$$\psi(z,t) = A\ f(z)e^{-i\omega t} \ . \quad (3)$$

The probability of finding the particle at z in dz is $|\psi(z,t)|^2$, which is independent of t. If the potential energy of the particle is constant, we have a "homogeneous medium", and f(z) is then a sinusoidal function of kz. Traveling waves have the form

$$\psi(z,t) = A\, e^{i(kz-\omega t)} \quad , \tag{4}$$

for waves traveling in the +z direction. Standing waves have the form

$$\psi(z,t) = A \sin(kz+\alpha) e^{-i\omega t} \quad . \tag{5}$$

If k^2 is negative, the waves have the form

$$\psi(z,t) = [A\, e^{-|k|z} + B\, e^{+|k|z}]\, e^{-i\omega t} \quad . \tag{6}$$

The dispersion relation for a particle in a region of constant potential energy V is obtained by substituting $E = \hbar\omega$ and $p = \hbar k$ (the Bohr frequency condition and the de Broglie wave-number relation) into the classical expression for the energy. For example, for non-relativistic electrons of mass m, the energy E, momentum p, and potential V are related classically by

$$E = \frac{p^2}{2m} + V \quad , \tag{7}$$

which gives the dispersion relation for the de Broglie waves to be

$$\hbar\omega = \frac{\hbar^2 k^2}{2m} + V \quad . \tag{8}$$

As long as $\hbar\omega$ is greater than V, k^2 is positive. Then $\psi(z,t)$ has a form like Eq. (4) or Eq. (5). For $\hbar\omega$ less than V, k^2 is negative. In classical mechanics this is "forbidden" in the sense that the kinetic energy $p^2/2m$ cannot be negative. But k^2 can be negative, because these are waves, not classical particles. When the wave enters a "classically forbidden" region (from a classically allowed one), the wave function is exponentially damped with increasing distance, as one sees from Eq. (6) with B = 0. (If the forbidden region is not very thick, both A and B are non-zero.)

As an example, we take the electron to be confined in a one-dimensional "box" extending from z = 0 to z = L. Inside the box, V is zero. Outside, V is huge and positive, so that k^2 is negative with huge magnitude. That means the wave is rapidly attenuated to zero outside the box. If we take the limiting case of V infinite, then the wave amplitude goes to zero at z = 0 and z = L and is zero everywhere outside the box. That is "just like" a violin string! And, just as for a continuous violin string, it turns out that the lowest mode for the electron de Broglie standing wave in the box has just one half wavelength between z = 0 and z = L. The next mode, number 2, has two half waves, etc; the allowed values of k are given by

$$k_n L = n\pi \, , \; n = 1,2, \ldots \quad . \tag{9}$$

The frequency depends on the wave number through Eq. (8), with V = 0:

$$E_n = \hbar\omega_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2(n\pi/L)^2}{2m}, \quad n = 1,2,3, \ldots \quad . \tag{10}$$

The physical meaning of the waves is that the space dependence

$$|\psi_n(z)|^2 = |A_n \sin k_n z|^2 \tag{11}$$

gives the probability that the electron is at any instant in some small interval at the position z, if it is in mode n. (In quantum mechanics one calls the modes by the name "stationary states".)

3. Phase and group velocities for de Broglie waves.

For a non-relativistic electron of energy E in a constant potential V, the dispersion relation is [see Eq. (8)]

$$\omega = \frac{\hbar k^2}{2m} + \frac{V}{\hbar} \quad . \tag{12}$$

The phase velocity is

$$v(k) = \frac{\omega}{k} = \frac{\hbar k}{2m} + \frac{V}{\hbar k} \quad . \tag{13}$$

The velocity of the classical particle is p/m, i.e., $\hbar k/m$. Thus Eq. (13) reads

$$v(k) = v(\text{particle}) + V/p(\text{particle}) \quad , \tag{14}$$

which is a peculiar relationship. Luckily, v(k) is not directly observable. The velocity of a classical particle is instead that of a "wave packet" (see Sec. 2.4) made up of several neighboring values of k, not just one value.

The velocity of propagation of a wave packet is given by the group velocity v_g. Thus, using Eq. (12),

$$v_g = \left(\frac{d\omega}{dk}\right)_o = \left(\frac{\hbar k}{m}\right)_o \quad , \tag{15}$$

where the subscript zero means one is to evaluate the derivative at the value of k at the center of the band Δk that forms the packet. Thus we see that v_g = v(particle) if we take the particle momentum to be $(\hbar k)_o$, corresponding to the center of the packet.

For a relativistic free particle the relation between energy, momentum, and rest mass m is given by

$$E^2 = (mc^2)^2 + (cp)^2 \quad , \tag{16}$$

which gives the dispersion relation (using $E = \hbar\omega$ and $p = \hbar k$, which are relativistically correct)

$$\hbar^2\omega^2 = (mc^2)^2 + (\hbar ck)^2 \quad . \tag{17}$$

The phase velocity $v = \omega/k$ has the value $v = \omega/k = E/p$, which is c^2/v(particle) and is thus greater than c. The group velocity is

$$v_g = \frac{d\omega}{dk} = \frac{c^2k}{\omega} = \frac{c^2p}{E} = v(\text{particle}) \quad . \tag{18}$$

The relations between the phase velocity, group velocity, and velocity of light are the same as those for radio waves in the ionosphere (Chap. 1, page 64), namely $v\, v_g = c^2$. That is because the dispersion relations are

similar [compare Eq. (17) above with Eq. (114), p. 41, Chap. 1].

4. Wave equations for de Broglie waves.

According to Eqs. (4) or (5), a harmonic (i.e., stationary state) de Broglie wave has the form in a region of constant potential

$$\psi(z,t) = e^{-i\omega t}\,(Ae^{ikz} + Be^{-ikz}) \quad . \tag{19}$$

Thus

$$\frac{\partial\psi(z,t)}{\partial t} = -i\omega\,\psi(z,t) \tag{20}$$

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = -\omega^2\psi\,(z,t) \tag{21}$$

$$\frac{\partial\psi(z,t)}{\partial z} = e^{-i\omega t}(ikAe^{ikz} - ikBe^{-ikz}) \tag{22}$$

$$\frac{\partial^2\psi(z,t)}{\partial z^2} = -k^2\psi(z,t) \tag{23}$$

For non-relativistic particles, the dispersion relation is given by [see Eq. (8)]

$$\hbar\omega = \frac{\hbar^2k^2}{2m} + V \quad . \tag{24}$$

Multiplying Eq. (20) by $i\hbar$ and using Eqs. (24) and (23), we obtain

$$i\hbar\frac{\partial\psi(z,t)}{\partial t} = -\frac{\hbar^2}{2m}\,\frac{\partial^2\psi(z,t)}{\partial z^2} + V\psi(z,t) \quad . \tag{25}$$

Equation (25) was derived using harmonic waves for a uniform potential, which gives a space dependence "sinusoidal" in kz. There is no reason why one can demand that Eq. (25) should hold when one has V = V(z), i.e., V is a function of position. But one can hope. That is what Schrödinger did. He guessed that perhaps Eq. (25) holds even for a V(z) which is not constant. Equation (25), with V = V(z), is called the Schrödinger equation (more precisely, the one-dimensional, time-dependent Schrödinger equation). It works, for atomic physics.

When we cannot neglect relativistic effects, we cannot use Eq. (24), nor Eq. (25). For free relativistic particles, the relativistic dispersion relation is [See Eq. (17)]

$$\hbar^2\omega^2 = \hbar^2c^2k^2 + (mc^2)^2 \quad . \tag{26}$$

Multiplying Eq. (26) by $-\hbar^{-2}\psi(z,t)$ and using Eqs. (21) and (23), we obtain

$$\frac{\partial^2\psi(z,t)}{\partial t^2} = c^2\frac{\partial^2\psi(z,t)}{\partial z^2} - \frac{(mc^2)}{\hbar^2}\psi(z,t) \quad . \tag{27}$$

Equation (27) is called the Klein-Gordon equation. Notice that if we set m = 0 we get the classical wave equation.

5. Three-dimensional classical wave equation with spherical symmetry.

The classical wave equation is given by

$$\begin{aligned}\frac{\partial^2\psi(x,y,z,t)}{\partial t^2} &= v^2\nabla^2\psi(x,y,z,t) \\ &= v^2\left[\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}\right]\end{aligned} \tag{28}$$

Let us assume $\psi(x,y,z,t)$ depends on x,y, and z only through r, where r is given by

$$r = (x^2 + y^2 + z^2)^{1/2} \quad . \tag{29}$$

[That is what we would expect for a wave function describing sound waves emitted by a pulsating (expanding and contracting) sphere centered at $r = 0$.] Thus we assume

$$\psi(x,y,z,t) = \psi(r,t) \quad .$$

Then

$$\frac{\partial\psi(r,t)}{\partial x} = \frac{\partial\psi(r,t)}{\partial r} \cdot \frac{\partial r}{\partial x} = \frac{x}{r} \frac{\partial\psi(r,t)}{\partial r} \quad , \tag{30}$$

where we used the fact that differentation of Eq. (29) with respect to x gives that $\partial r/\partial x$ equals x/r. Regarding the right-hand side of Eq. (30) as a product of x and $r^{-1}\,\partial\psi/\partial r$, and differentiating with respect to x, we find

$$\frac{\partial^2\psi(r,t)}{\partial x^2} = \frac{\partial}{\partial x}\left\{x\, r^{-1} \frac{\partial\psi}{\partial r}\right\} = 1 \cdot r^{-1} \cdot \frac{\partial\psi}{\partial r} + x \cdot \frac{x}{r} \cdot \frac{\partial}{\partial r}\left\{r^{-1} \frac{\partial\psi}{\partial r}\right\} \quad , \tag{31}$$

where the factor x/r in the last term is $\partial r/\partial x$. The Laplacian of ψ (i.e., $\nabla^2\psi$) is obtained by adding Eq. (31) to the analogous expressions for $\partial^2\psi/\partial y^2$ and $\partial^2\psi/\partial z^2$. Because of the spherical symmetry, these may be obtained from Eq. (31) by merely replacing x by y or z. The Laplacian is

thus given by

$$\nabla^2\psi(r,t) = 3r^{-1}\frac{\partial\psi}{\partial r} + \frac{(x^2+y^2+z^2)}{r}\left\{-r^{-2}\frac{\partial\psi}{\partial r} + r^{-1}\frac{\partial^2\psi}{\partial r^2}\right\}$$

$$= 2r^{-1}\frac{\partial\psi}{\partial r} + \frac{\partial^2\psi}{\partial r^2} = r^{-1}\left\{2\frac{\partial\psi}{\partial r} + r\frac{\partial^2\psi}{\partial r^2}\right\}$$

$$= r^{-1}\left\{\frac{\partial\psi}{\partial r} + \frac{\partial}{\partial r}\left[r\frac{\partial\psi}{\partial r}\right]\right\} = r^{-1}\frac{\partial}{\partial r}\left\{\psi + r\frac{\partial\psi}{\partial r}\right\}$$

$$= r^{-1}\frac{\partial}{\partial r}\left\{\frac{\partial}{\partial r}(r\psi)\right\} = r^{-1}\frac{\partial^2}{\partial r^2}(r\psi) \quad . \qquad (32)$$

Thus the wave equation, Eq. (28), becomes

$$\frac{\partial^2\psi(r,t)}{\partial t^2} = v^2 r^{-1}\frac{\partial^2(r\psi)}{\partial r^2} \quad . \qquad (33)$$

We multiply Eq. (33) by r and find

$$\frac{\partial^2}{\partial t^2}[r\psi(r,t)] = v^2\frac{\partial^2[r\psi(r,t)]}{\partial r^2} \qquad (34)$$

But Eq. (34) looks just like our familiar one-dimensional wave equation,

$$\frac{\partial^2\varphi(z,t)}{\partial t^2} = v^2\frac{\partial^2\varphi(z,t)}{\partial z^2} \quad , \qquad (35)$$

except that instead of z we have r and instead of φ we have $r\psi$. Thus let us define

$$\varphi(r,t) = r\psi(r,t) \quad . \tag{36}$$

Then by analogy with the general solution of Eq. (35), which we know is an arbitrary superposition of traveling waves of the form

$$\varphi(z,t) = F(z-vt) + G(z+vt) \ , \tag{37}$$

we must have

$$\varphi(r,t) = F(r-vt) + G(r+vt) \ . \tag{38}$$

The traveling wave $r^{-1} F(r-vt)$ is called an outgoing spherical wave; similarly $r^{-1} G(r+vt)$ is an incoming spherical wave:

$$\psi_{out}(r,t) = \frac{F(r-vt)}{r} \ , \tag{39}$$

$$\psi_{in}(r,t) = \frac{G(r+vt)}{r} \ . \tag{40}$$

If the source of waves is a pure harmonic oscillation, we get harmonic spherical outgoing waves (or incoming waves -- the source could be the pulsating walls of a hollow sphere). The general solution for a given ω is thus

$$\psi(r,t) = \frac{A}{r} \cos(\omega t - kr + \phi_1) + \frac{B}{r} \cos(\omega t + kr + \phi_2) \ . \tag{41}$$

We may worry about the fact that $\psi(r,t)$ as given by Eqs. (39), (40), and (41) becomes infinite at $r = 0$. That is no real worry, because if there is a transmitter near the origin it has to occupy a finite space, and thus the coupling between the transmitter and the wave occurs at a finite value of r .

6. Reflection of de Broglie waves at discontinuities in potential energy.

First you should review Advanced Topic 2, where we gave the dispersion relation for de Broglie waves. According to Eq. (8), the dispersion relation is

$$\hbar^2 k^2 = 2m(\hbar\omega - V) \quad . \tag{42}$$

Suppose that medium 1 extends from $z = -\infty$ to $z = 0$ and has $\hbar\omega - V_1$ positive, with V_1 constant. Then medium 1 is dispersive. An incident traveling wave propagating in the + z direction has wave function $\psi_{inc}(z,t)$ given by

$$\psi_{inc}(z,t) = A\, e^{-i\omega t}\, e^{i\, k_1\, z} \quad . \tag{43}$$

At $z = 0$ the potential suddenly changes from V_1 to V_2. The transmitted wave function has the form $\psi_{tr}(z,t)$ given by

$$\psi_{tr}(z,t) = B\, e^{-i\omega t}\, e^{i\, k_2\, z} \tag{44}$$

where k_2 is $+ i|k_2|$ if k_2^2 is negative. If medium 2 is a reactive region, then $|k_2|$ is given by [using Eq. (42)]

$$\hbar|k_2| = [2m(V_2 - \hbar\omega)]^{1/2} \quad . \tag{45}$$

The reflection coefficient for either case (dispersive or reactive medium 2) is obtained by considering what happens at a discontinuity in V. Now, $\psi^*(z,t)\,\psi(z,t)$ represents the probability of finding the electron at (z,t). This quantity should not jump discontinuously at a boundary. Therefore

$\psi(z,t)$ should be continuous. Also, the quantity $\psi^*(z,t)\,\partial\psi(z,t)/\partial z$ plus its complex conjugate turns out to be proportional to the particle flux. That should also be continuous. Therefore $\partial\psi(z,t)/\partial z$ should be continuous along with $\psi(z,t)$. Writing $\psi_1(z,t)$ for the superposition of the incident and reflected traveling waves in medium 1, we have

$$\psi_1(z,t) = Ae^{-i\omega t}\left[e^{i k_1 z} + R_\psi e^{-ik_1 z}\right] \quad . \tag{46}$$

In medium 2 we have the transmitted wave $\psi_2(z,t)$ given by

$$\psi_2(z,t) = Be^{-i\omega t} e^{i k_2 z} \quad . \tag{47}$$

The boundary conditions at $z = 0$ are that ψ_1 and ψ_2 be equal and that $\partial\psi_1/\partial z$ and $\partial\psi_2/\partial z$ be equal. These give [using Eqs. (46) and (47)]

$$B = A(1 + R_\psi) \quad , \tag{48}$$

$$Bk_2 = A\,k_1(1 - R_\psi) \quad . \tag{49}$$

These equations may be solved for R_ψ and B/A:

$$R_\psi = \frac{k_1 - k_2}{k_1 + k_2} \quad , \tag{50}$$

$$T_\psi \equiv \frac{B}{A} = 1 + R_\psi = \frac{2\,k_1}{k_1 + k_2} \quad . \tag{51}$$

This result for R_ψ has exactly the same form as that which we found for the reflection coefficient R_E of the electric field $E_x(z,t)$ for the case of a plane wave in a dispersive medium incident on a discontinuity in the dielectric constant ϵ (but where the magnetic permeability μ is the same in both media). [See Eq. (137), Sec. 3.4.] The reflection coefficient R_ψ for the wave function $\psi(z,t)$ of a de Broglie wave thus has the same form as that for the electric field in an electromagnetic plane wave (with constant magnetic permeability). The space derivative $\partial\psi(z,t)/\partial z$ has a reflection coefficient that equals $-R_\psi$. This has the same form as the reflection coefficient of the magnetic field in an electromagnetic plane wave. Of course these are just analogies, but they are interesting.

Now suppose that medium 2 is reactive. Then we have

$$k_2 = +i|k_2| \quad . \tag{52}$$

Then R_ψ has an absolute magnitude of 1 (as shown in Sec. 3.7) and thus may be written in the form

$$R_\psi = -1\, e^{i\delta} \quad . \tag{53}$$

The wave function in medium 1 is therefore given by

$$\begin{aligned}
\psi_1(z,t) &= A\, e^{-i\omega t}\left[e^{ik_1 z} - e^{i\delta}\, e^{-ik_1 z}\right] \\
&= A\, e^{-i\omega t}\, e^{i\frac{1}{2}\delta}\left[e^{i(k_1 z - \frac{1}{2}\delta)} - e^{-i(k_1 z - \frac{1}{2}\delta)}\right] \\
\psi_1(z) &= A\, e^{-i(\omega t - \frac{1}{2}\delta)}\, 2i \sin\,(k_1 z - \tfrac{1}{2}\delta) \quad .
\end{aligned} \tag{54}$$

The particle density in medium 1 is equal to $|\psi_1(z,t)^2|$:

$$\text{particle density} = 4A^2 \sin^2 (k_1 z - \tfrac{1}{2}\delta) \quad . \tag{55}$$

The transmitted wave function is given by

$$\psi_{tr}(z,t) = Be^{-i\omega t} e^{-|k_2|z} = A e^{-i\omega t} (1 - e^{i\delta}) e^{-|k_2|z} \tag{56}$$

$$= A e^{-i\omega t} e^{+i\frac{1}{2}\delta} \left(e^{-i\frac{1}{2}\delta} - e^{i\frac{1}{2}\delta}\right) e^{-|k_2|z}$$

$$= -2iAe^{-i(\omega t - \frac{1}{2}\delta)} \sin \tfrac{1}{2}\delta \, e^{-|k_2|z} \tag{57}$$

The transmitted particle density is given by $|\psi_{tr}|^2$:

$$\text{transmitted particle density} = 4A^2 \sin^2 \tfrac{1}{2}\delta \, e^{-2|k_2|z} \quad . \tag{58}$$

The most important factor in Eq. (58) is the exponential attenuation factor $e^{-2|k_2|z}$.

Barrier penetration. If the reactive region (number 2) [the region where V_2 exceeds $\hbar\omega$] does not extend to $z = +\infty$ but only to $z = +L$, and a dispersive region (number 3) extends from $z = L$ to $z = +\infty$, then the wave function in region 2 must include a term in $e^{+|k_2|z}$ as well as the term in $e^{-|k_2|z}$. But if L is large compared with the mean attenuation distance $|k_2|^{-1}$, then it turns out that we can neglect this term. Then the wave transmitted into medium 3 is proportional to the attenuation factor

$$\text{attenuation factor} = e^{-2|k_2|L} \quad . \tag{59}$$

where $|k_2|$ is given by Eq. (45). Barrier penetration factors like that given in Eq. (59) play an important role in quantum mechanics. They are one of the extremely "non-classical" results of quantum mechanics, where "non-classical" means "not to be understood on the basis of Newtonian mechanics of particles." According to Newtonian mechanics it makes no sense to speak of a particle with its total (kinetic plus potential) energy E less than its potential energy V, because then its kinetic energy is negative. According to quantum mechanics, one then has a reactive medium for the de Broglie waves associated with the particle. The particle can be found there and can "leak" through a classically forbidden region by barrier penetration.

There is something of an analogy between classical mechanics and geometrical optics. In geometrical optics, one says that Snell's law "forbids" the light to be transmitted from glass to air if the angle of incidence exceeds the critical angle (see page 92, Chap. 3). The wave theory of optics tells us that there is electromagnetic energy density (light) in the "forbidden" region. It falls to zero exponentially, but light can be transmitted across the optical barrier into a second piece of glass if the barrier is not infinitely thick.

7. Solution of driven oscillator problem by Green's function method. (See Prob. 13, Chap. 2.)

Let x(t) represent the displacement from equilibrium of an undamped harmonic oscillator with natural oscillation frequency ω and mass m that

is driven by a force F(t). The problem is to find x(t), given the starting conditions that x and dx/dt are both zero at time zero.

The differential equation for x(t) is obtained from Newton's second law:

$$\frac{md^2x}{dt^2} = -C\,x + F(t) = -m\,\omega^2 x(t) + F(t) \qquad (60)$$

In preparation for solving this equation for an arbitrary function F(t), we consider its solution for a very special force which we shall call $f_1(t)$. The force $f_1(t)$ is an "impulse" that occurs at time t_1. This means that $f_1(t)$ is zero for all times t except during a small interval Δt_1 centered at t_1. During this small interval, $f_1(t)$ has a constant value, $f_1(t_1)$. The product $f_1(t_1)\Delta t_1$ is the momentum p_1 delivered to the mass. The solution of Eq. (60) with $F(t) = f_1(t)$ and with the conditions that x and dx/dt be zero at $t = 0$ is easily obtained. For times less than t_1, the mass remains at rest. At t_1 it suddenly acquires momentum p_1, i.e., velocity p_1/m. After that it oscillates freely. Thus the solution is given by

$$\begin{aligned} \frac{dx}{dt} &= 0 \text{ for } -\infty < t < t_1 \\ &= \frac{p_1}{m} \cos \omega(t - t_1) \text{ for } t_1 \leq t < \infty \end{aligned} \qquad (61)$$

Then x(t) is obtained by integrating dx/dt with respect to t:

$$\begin{aligned} x &= 0 \text{ for } -\infty < t < t_1 \\ &= \frac{p_1}{m\omega} \sin \omega(t - t_1) \text{ for } t_1 \leq t < \infty \end{aligned} \qquad (62)$$

Now that we have seen what happens when a single impulsive force is applied, let us consider the case where there are two impulses, one at time t_1 and one at time t_2. The force function F(t) is then zero except in a small interval Δt_1 near time t_1 and another small interval Δt_2 near t_2. We can think of F(t) as the superposition of the two impulsive forces. (Only one of the impulses is turned on at a time, i.e., t_2 is not equal to t_1.) Because of the linearity of Eq. (60), the motion x(t) induced by a force F(t) that is the superposition of two impulses $f_1(t)$ and $f_2(t)$ is the superposition of the motions that would result from each individual impulse if each were present by itself. (You may easily prove that statement.) Thus for the case of two impulses we have

$$
\begin{aligned}
x &= 0 \text{ for } -\infty \leq t < t_1 \\
m\omega x &= p_1 \sin \omega(t - t_1) \text{ for } t_1 \leq t < t_2 \\
&= p_1 \sin \omega(t - t_1) + p_2 \sin \omega(t - t_2) \\
&\qquad \text{for } t_2 \leq t < +\infty \quad . \\
&= f_1(t_1)\Delta t_1 \sin \omega(t - t_1) + f_2(t_2)\Delta t_2 \sin \omega(t - t_2) \, , \\
&\quad t_2 \leq t < \infty \quad . \qquad (63)
\end{aligned}
$$

Now comes the beautiful trick. Any force F(t) that is a continuous function of time and is applied for a time interval from 0 to t can be thought of as a superposition of impulses. The first impulse is applied between time zero and time Δt_1; the second is applied between Δt_1 and $\Delta t_1 + \Delta t_2$, etc.

Thus the motion $x(t)$ at time t is the superposition of the motions induced by all the impulses between time zero and t. Generalizing Eq. (63), we obtain

$$m\,\omega\,x(t) = p_1 \sin\omega(t - t_1) + p_2 \sin\omega(t - t_2) + p_3 \sin\omega(t - t_3) + \ldots$$

$$= f_1(t_1)\Delta t_1 \sin\omega(t - t_1) + f_2(t_2)\,\Delta t_2 \sin\omega(t - t_2) + \ldots \quad . \qquad (64)$$

Now we go to the limit in which all the time intervals Δt_1, Δt_2, ... are so small that we may write Eq. (64) as an integral. The final result is

$$m\,\omega\,x(t) = \int_0^t F(t') \sin\omega(t - t')\,dt' \quad , \qquad (65)$$

where $F(t')$ is the force at time t'. Equation (65) satisfies the differential equation Eq. (60) and the boundary conditions that x and dx/dt are zero at $t = 0$. It is the general solution and holds for any reasonable $F(t)$.

Equation (65) is an integral equation. It relates $x(t)$ to $x(0)$ by means of an integral over all the intervening times t'. The integrand per unit force that "connects" t' and t, i.e., $\sin\omega(t-t')$, is called the "Green's function." Notice that Eq. (62), the solution for a single impulsive force, could be written in the form

$$m\,\omega\,x(t) = F(t')\,\Delta t' \sin\omega(t - t') \quad , \qquad (66)$$

and Eq. (65) is the superposition of many such contributions.

8. Angular momentum carried by electromagnetic plane waves.

In Sec. 3.5 we found the radiation emitted by an oscillating sheet of charge that covers the entire xy plane at $z = 0$. Suppose that these charges are driven in circles with the sense of rotation along $+\hat{z}$. (We use a right-hand rule to define the sense of rotation.) The driving force will have to exert a torque to keep them moving steadily in a circle. Therefore the driving force loses angular momentum (which it may have stored in a big rotating wheel). The circularly polarized radiation carries this angular momentum. If it is absorbed by a perfectly terminating resistive load, it will deliver the same angular momentum to the load that it acquired from the driving force.

We wish to find the relation between the energy radiated and the angular momentum radiated for circularly polarized electromagnetic waves. Suppose that the charges all go around in circles in little frictionless tubes of constant radius. Suppose that at $t = 0$ the driving force is removed, and the charges slowly lose their energy and angular momentum to the radiated circularly polarized light.

At any instant the kinetic energy of a charge of mass M is

$$E = 1/2\, M\, v^2 = 1/2\, M\, r^2\, \omega^2 \quad , \tag{67}$$

where r is the radius of the tube and ω is the angular frequency of rotation.

The angular momentum of the charge is

$$J = Mvr = Mr^2 \omega \quad . \tag{68}$$

During a certain time interval, the charges lose energy and angular momentum. The radius is fixed, so the frequency must slow down. Thus as ω decreases by $d\omega$, the energy decreases by dE and the angular momentum decreases by dJ. According to Eqs. (67) and (68), the ratio of these is

$$\frac{dE}{dJ} = \frac{Mr^2 \omega \, d\omega}{Mr^2 d\omega} = \omega \quad . \tag{69}$$

Thus, according to Eq. (69), whenever the source loses energy ΔE, it loses angular momentum $\Delta E/\omega$.

This result still holds in quantum theory. A circularly polarized photon of energy $\hbar\omega$ has angular momentum $\pm \hbar$ along the z (propagation) direction.

Appendix

Taylor's Series

Assume that f(x) can be written as an infinite series of the form

$$f(x) = c_o + c_1(x-x_o) + c_2(x-x_o)^2 + c_3(x-x_o)^3 + \cdots , \qquad (1)$$

where the c's are constants. Then f(x) is said to be expressed as "an expansion at point x_o." To find c_o we set $x = x_o$. All terms on the right-hand side vanish except the first. Thus $c_o = f(x_o)$. To find c_1, we differentiate Eq. (1) once with respect to x and then set $x = x_o$. All terms vanish except the c_1 term; thus we find $c_1 = (df/dx)_o$, where the subscript zero means that df/dx is evaluated at $x = x_o$. Similarly,

$$(d^m f/dx^m)_o = m!\, c_m , \qquad (2)$$

and Eq. (1) becomes

$$f(x) = f(x_o) + (x-x_o)(df/dx)_o + \frac{(x-x_o)^2}{2!}(d^2f/dx^2)_o + \frac{(x-x_o)^3}{3!}(d^3f/dx^3)_o + \cdots \qquad (3)$$

Commonly-used series

sinx and cosx. Use d(sinx)/dx = cosx, d(cosx)/dx = -sinx, cos(0) = 1, sin(0) = 0, and $x_o = 0$ in Eq. (3) to obtain

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots , \qquad (4)$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots . \qquad (5)$$

Exponential e^{ax}. Use $d(e^{ax})/dx = ae^{ax}$, $e^{a(0)} = 1$, and $x_o = 0$ in Eq. (3) to obtain

$$e^{ax} = 1 + ax + \frac{a^2x^2}{2!} + \frac{a^3x^3}{3!} + \frac{a^4x^4}{4!} + \cdots . \qquad (6)$$

sinh x and cosh x. These functions may be defined by $d(\sinh x)/dx = \cosh x$, $d(\cosh x)/dx = \sinh x$, $\sinh(0) = 0$, $\cosh(0) = 1$, and Eq. (3) with $x_o = 0$, to give

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots , \tag{7}$$

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots . \tag{8}$$

Relationships involving the exponential. If we set $a = +1$ in Eq. (6) and compare with Eqs. (7) and (8), and if we then do the same for $a = -1$, we find

$$e^x = \cosh x + \sinh x , \tag{9}$$

$$e^{-x} = \cosh x - \sinh x , \tag{10}$$

which may be solved to obtain

$$\cosh x = \frac{e^x + e^{-x}}{2} , \tag{11}$$

$$\sinh x = \frac{e^x - e^{-x}}{2} . \tag{12}$$

If we set $a = +i \equiv +\sqrt{-1}$ in Eq. (6), we obtain

$$e^{ix} = 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} + \cdots . \tag{13}$$

Similarly, if we set $a = -i$ in Eq. (6), we obtain

$$e^{-ix} = 1 - ix - \frac{x^2}{2!} + \frac{ix^3}{3!} + \frac{x^4}{4!} - \frac{ix^5}{5!} - \frac{x^6}{6!} + \cdots . \tag{14}$$

By adding or subtracting Eqs. (13) and (14) and comparing the results with Eqs. (4) and (5), we obtain

$$\frac{e^{ix} + e^{-ix}}{2} = \cos x, \tag{15}$$

$$\frac{e^{ix} - e^{-ix}}{2i} = \sin x, \tag{16}$$

which may be solved to give

$$e^{ix} = \cos x + i\sin x \,, \tag{17}$$

$$e^{-ix} = \cos x - i\sin x \,. \tag{18}$$

tan x. Use $\tan x \equiv \sin x/\cos x$, $d(\sin x)/dx = \cos x$, $d(\cos x)/dx = -\sin x$, to obtain $d(\tan x)/dx = (\cos x)^{-2}$, $d^2(\tan x)/dx^2 = 2\sin x(\cos x)^{-3}$, $d^3(\tan x)/dx^3 = 2(\cos x)^{-2} - 6\sin^2 x(\cos x)^{-4}$, etc. Then use $x_o = 0$ in Eq. (3) to obtain

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \cdots \,. \tag{19}$$

Binomial series $(1 + x)^n$. Use $d(1+x)^n/dx = n(1+x)^{n-1}$, $d^2(1+x)^n/dx^2 = n(n-1)(1+x)^{n-2}$, $d^3(1+x)^n/dx^3 = n(n-1)(n-2)(1+x)^{n-3}$, etc. and use Eq. (3) with $x_o = 0$ to obtain

$$(1+x)^n = 1 + nx + \frac{n(n-1)x^2}{2!} + \frac{n(n-1)(n-2)x^3}{3!} + \cdots \,. \tag{20}$$

Eq. (20) holds for any n, positive or negative, and any x, positive or negative, as long as x satisfies the relation $x^2 < 1$.

Superposition of harmonic functions

The following superpositions of N harmonic functions are encountered in wave phenomena:

$$u(t) = \cos \omega_1 t + \cos(\omega_1+\alpha)t + \cos(\omega_1+2\alpha)t + \cdots + \cos[\omega_1+(N-1)\alpha]t; \tag{21}$$

$$u(z) = \cos kz + \cos(kz + \beta) + \cos(kz+2\beta) + \cdots + \cos [kz+(N-1)\beta] \tag{22}$$

These are both of the form

$$u = \cos \theta_1 + \cos(\theta_1 + \gamma) + \cos (\theta_1 + 2\gamma) + \cdots + \cos [\theta_1 +(N-1)\gamma] \tag{23}$$

We wish to find a convenient expression for Eq. (23). We note that u can be written as the real part of v, where

$$v = e^{i\theta_1} + e^{i(\theta_1+\gamma)} + e^{i(\theta_1+2\gamma)} + \cdots + e^{i[\theta_1+(N-1)\gamma]} = e^{i\theta_1} S \tag{24}$$

where S is a geometric series of N terms, given by

$$S = 1 + a + a^2 + a^3 + \cdots + a^{N-1}, \quad \text{with } a = e^{i\gamma} . \tag{25}$$

Multiply S by a. Then subtract S from aS, term by term, to get

$$aS - S = a^N - 1, \tag{26}$$

i.e.,

$$S = \frac{a^N-1}{a-1} = \frac{e^{iN\gamma} - 1}{e^{i\gamma} - 1} = \frac{e^{\frac{1}{2}iN\gamma}}{e^{\frac{1}{2}i\gamma}} \frac{(e^{\frac{1}{2}iN\gamma} - e^{-\frac{1}{2}iN\gamma})}{(e^{\frac{1}{2}i\gamma} - e^{-\frac{1}{2}i\gamma})} = e^{\frac{1}{2}i(N-1)\gamma} \frac{\sin \frac{1}{2} N\gamma}{\sin \frac{1}{2} \gamma} \tag{27}$$

where we used Appendix Eq. (16) in the last step. Inserting Eq. (27) into Eq. (24) gives

$$v = e^{i[\theta_1 + \frac{1}{2}(N-1)\gamma]} \frac{\sin \frac{1}{2} N\gamma}{\sin \frac{1}{2}\gamma} . \tag{28}$$

Taking the real part, we obtain the desired result,

$$u = \cos [\theta_1 + \tfrac{1}{2}(N-1)\gamma] \frac{\sin \frac{1}{2} N \gamma}{\sin \frac{1}{2} \gamma} . \tag{29}$$

The result, Eq. (29), can be put in another useful form. In Eq. (23) θ_1 is the argument of the first term, and the argument of the last term, which we shall call θ_2, is given by

$$\theta_1 + (N-1)\ \gamma \equiv \theta_2 . \tag{30}$$

The average of the first and last arguments θ_1 and θ_2 is then

$$\theta_{av} \equiv \tfrac{1}{2} (\theta_1+\theta_2) = \tfrac{1}{2}\theta_1 + \tfrac{1}{2}\theta_1 + \tfrac{1}{2}(N-1)\ \gamma . \tag{31}$$

Thus the first factor in Eq. (29) is just cos θ_{av}. Using this and the fact that γ is $(\theta_2-\theta_1)/(N-1)$ [according to Eq. (30)], we write Eq. (29) in the form

$$u = \cos \theta_{av} \frac{\sin [\frac{1}{2}N(\theta_2-\theta_1)/(N-1)]}{\sin [\frac{1}{2}(\theta_2-\theta_1)/(N-1)]} . \tag{32}$$

Equation (29) emphasizes the increment γ for successive terms in the sum, Eq. (23). Equation(32)is equivalent to Eq.(29)but emphasizes the first and last contributions θ_1 and θ_2, and their average. Note that cos θ_{av} is a harmonic oscillation of the same form as each of the terms in the superposition, Eq. (23); but instead of amplitude unity, it has amplitude $A(\theta_1,\theta_2,N)$ given by the expression

$$A(\theta_1,\theta_2,N) = \frac{\sin[\frac{1}{2}N(\theta_2-\theta_1)/(N-1)]}{\sin[\frac{1}{2}(\theta_2-\theta_1)/(N-1)]} \quad . \tag{33}$$

The most compact expression of our result is then

$$u = A(\theta_1,\theta_2,N) \cos \theta_{av} \quad . \tag{34}$$

The case N = 2 corresponds for oscillations in time [Eq. (21)] to the phenomenon of "beats" and for oscillations in space [Eq. (22)] to the two-slit interference pattern. For oscillations in time, larger N gives "multiple beats," which produce a "pulse" behavior in u(t) in the limit as $N \to \infty$. (See Chap. 2, Sec. 2.4.) For oscillations in space, larger N gives the multiple-slit interference pattern, and the limit of $N \to \infty$ gives the single-slit diffraction pattern due to a single slit that is many wavelengths wide. (See Chap. 5, Sec. 5.6.)

Vector identities.

We shall use A, B, C to stand for scalar functions of x, y, and z, i.e., A(x,y,z), B(x,y,z), and C(x,y,z). Similarly $\underset{\sim}{A}$, $\underset{\sim}{B}$, and $\underset{\sim}{C}$ stand for vector functions of x, y, and z. Thus $\underset{\sim}{A}$ means $\hat{x}\, A_x(x,y,z)$ and $\hat{y}\, A_y(x,y,z) + \hat{z}\, A_z(x,y,$ where $\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit vectors. We want to learn how to work with the object $\underset{\sim}{\nabla}$ (pronounced "del") that is both a vector and a "take the derivative" operator. The trick is to write equations involving del so as to satisfy both its vector aspects and its "take the derivative" aspects. For example, ir

$$\underset{\sim}{\nabla}(A\,B) = (\underset{\sim}{\nabla}\,A)\,B + A(\underset{\sim}{\nabla}\,B) = B\,\underset{\sim}{\nabla}\,A + A\,\underset{\sim}{\nabla}\,B\,, \tag{35}$$

the first equality comes from the rule for differentiation of a product: first B is held constant, then A. The second equality gets rid of the parentheses, because, by convention, del differentiates only what is to the right of it. We can temporarily symbolize this by writing del as $\underset{\sim}{\nabla}_a$ when it is to operate only on $\underset{\sim}{A}$ (or A) and $\underset{\sim}{\nabla}_b$ when it is to operate on B (or $\underset{\sim}{B}$). In that way we take care of the product differentiation rule by adding subscripts. Then we move the operators and vectors around so that quantities not to be differentiated are"safely"on the left side of del, making sure that we satisfy the vector rules as we do so. Finally we erase the subscripts a and b. Thus

$$\underset{\sim}{\nabla}(AB) = \underset{\sim}{\nabla}_a(AB) + \underset{\sim}{\nabla}_b(AB) = B\underset{\sim}{\nabla}_a A + A\underset{\sim}{\nabla}_b B = B\underset{\sim}{\nabla}A + A\underset{\sim}{\nabla}B\,. \tag{36}$$

Similarly

$$\underset{\sim}{\nabla}\times(\underset{\sim}{A}\underset{\sim}{B}) = \underset{\sim}{\nabla}_a\times(\underset{\sim}{A}\underset{\sim}{B}) + \underset{\sim}{\nabla}_b\times(\underset{\sim}{A}\underset{\sim}{B}) = B\underset{\sim}{\nabla}_a\times\underset{\sim}{A} - \underset{\sim}{A}\times\underset{\sim}{\nabla}_b B = B\underset{\sim}{\nabla}\times\underset{\sim}{A} - \underset{\sim}{A}\times\underset{\sim}{\nabla}B \tag{37}$$

After some practice, you don't need to write the intermediate steps.

Now we wish to find an identity for $\underset{\sim}{\nabla}\times(\underset{\sim}{\nabla}\times\underset{\sim}{C})$. We assume you know the identity

$$\underset{\sim}{A}\times(\underset{\sim}{B}\times\underset{\sim}{C}) = \underset{\sim}{B}(\underset{\sim}{A}\cdot\underset{\sim}{C}) - \underset{\sim}{C}(\underset{\sim}{A}\cdot\underset{\sim}{B}) \quad \text{("back minus cab" rule)} \tag{38a}$$

$$= \underset{\sim}{B}(\underset{\sim}{A}\cdot\underset{\sim}{C}) - (\underset{\sim}{A}\cdot\underset{\sim}{B})\underset{\sim}{C}\,. \tag{38b}$$

We can use this rule, substituting del for $\underset{\sim}{A}$ and del for $\underset{\sim}{B}$. We must be careful to keep both dels on the left side of $\underset{\sim}{C}$, because both dels are supposed to differentiate $\underset{\sim}{C}$. Thus we cannot use Eq. (38a). We must use Eq. (38b). Then we get

$$\underset{\sim}{\nabla}\times(\underset{\sim}{\nabla}\times\underset{\sim}{C}) = \underset{\sim}{\nabla}(\underset{\sim}{\nabla}\cdot\underset{\sim}{C}) - (\underset{\sim}{\nabla}\cdot\underset{\sim}{\nabla})\underset{\sim}{C}\,. \tag{39}$$

In terms of x, y, and z components, Eq. (39) means

$$[\underset{\sim}{\nabla} \times (\underset{\sim}{\nabla} \times \underset{\sim}{C})]_x = \frac{\partial(\underset{\sim}{\nabla} \cdot \underset{\sim}{C})}{\partial x} - \nabla^2 C_x \ , \tag{40}$$

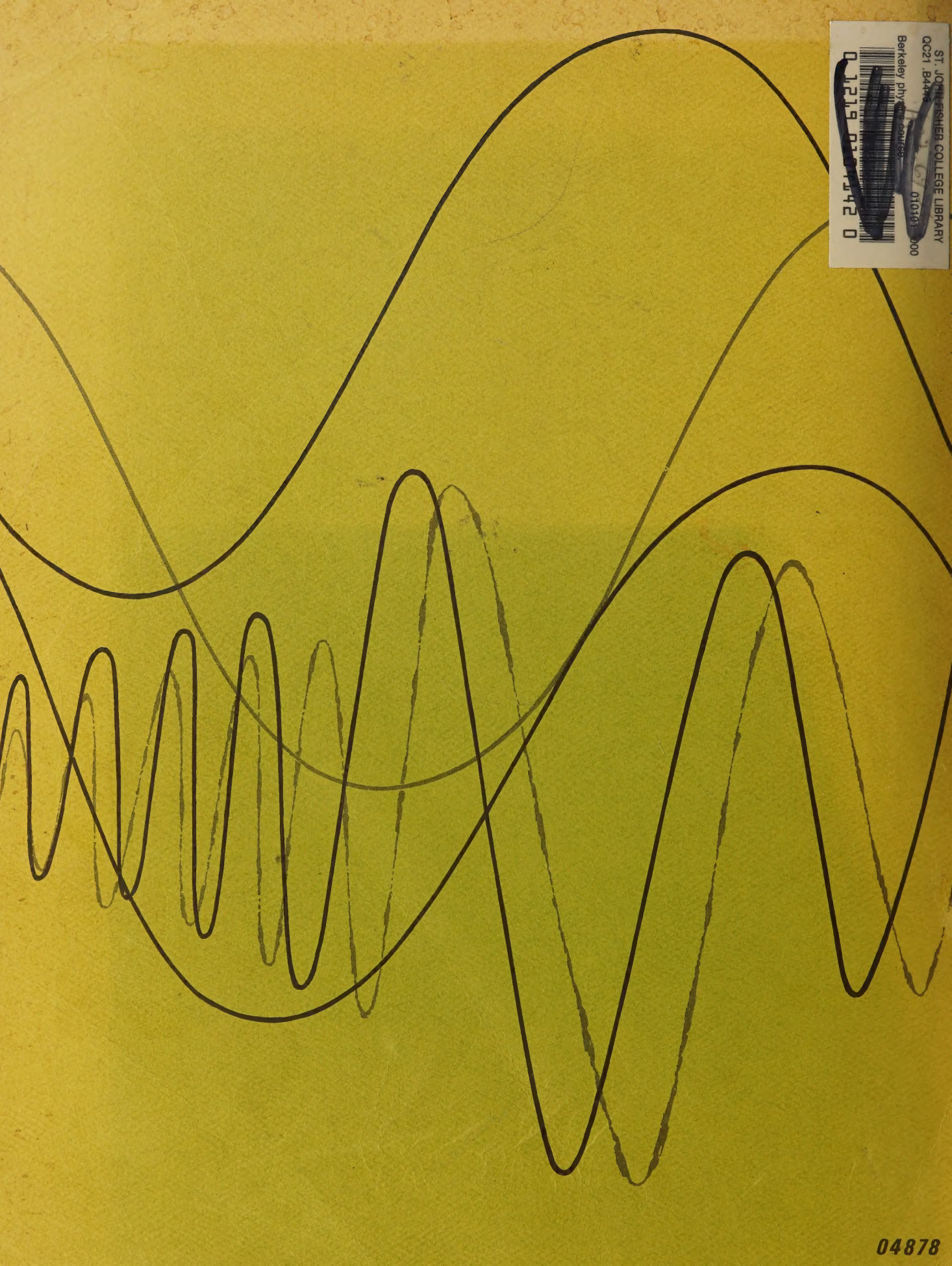
with similar expressions for y and z, and where

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \ . \tag{41}$$

ST. JOHN FISHER COLLEGE LIBRARY
QC21 .B4445
AAL6956
010101 000
Berkeley physics course.
0 1219 0104142 0

ST. JOHN FISHER COLLEGE LIBRARY

QC
21
.B4445
vol.3



ST. JOHN FISHER COLLEGE LIBRARY
QC21 .B4445
Berkeley physics course
010101 000
0 1219 0[illegible]42 0
04878